

Reading and Mathematics Intervention Programmes with  
Parents of Nursery Children in Disadvantaged Areas:  
a Psychological and Methodological Study

Walter Barker

Submitted in fulfilment of the requirements for the degree of  
Doctor of Philosophy in the Faculty of Education

Department of Child Development and Educational Psychology  
Institute of Education

University of London

1983

**THIS THESIS CONTAINS  
MICROFILMS THAT COULD  
NOT BE DIGITISED.**

**PLEASE CONTACT THE  
UNIVERSITY IF YOU WISH TO  
SEE THIS MATERIAL.**



### Acknowledgements

The author wishes to thank in particular:

Angela Hobsbaum, Lecturer, Department of Child Development and Educational Psychology, Institute of Education, University of London

and

Dr. John Versey, Senior Lecturer, Department of Child Development and Educational Psychology, Institute of Education, University of London

for the invaluable and unfailing help which they have each given to him, helping to direct his early studies in the fields of child development, educational psychology and applied statistics, and ultimately offering sensitive guidance and much encouragement in the study reported here

Olive Barker, infant teacher, Inner London Education Authority

without whose endless encouragement and total support during his years as a mature student, this study could not have been attempted

Colin Chalmers, Senior Lecturer, Department of Mathematics and Statistics, Birkbeck College, University of London

whose profound insights and willingness to encourage the exploration of new ideas made possible the refinement of the principal statistical technique developed in this study

More than 200 children, most of them living in poverty, whose friendliness and willingness to be assessed on a variety of challenging tasks, on three separate occasions, were crucial to the study

The parents of those children, and especially the 100 who attended the study programmes, whose sensitive concern for their children's future provided not only the justification but also many of the ideas noted in the study

The staff at the six schools where the intervention programmes were undertaken, and at two pilot schools: in particular the heads, who facilitated the study throughout its long presence in their schools, the nursery and infant teachers, who went to a great deal of trouble to enable the organisation of the assessment sessions and programme meetings, and the nursery assistants, those quite unassuming people whose insights and help also contributed much to the study

The Inner London Education Authority, whose remarkable openness to responsible inquiry is but one of many tributes which can be paid to this body, and in particular to its Research and Statistics Department

Dr. Rick Vinod, a statistician at American Telephone and Telegraph, who generously offered his ridge regression algorithm to this author, enabling the development of what may prove to be one of the most powerful and reliable forms of regression available today

Professor Colin Hindley, Department of Child Development and Educational Psychology, Institute of Education, and the many other members of the Department's staff whose interest and ideas contributed to earlier work and indirectly to the present study

Professor Harvey Goldstein, Tony Halil, Linda Noble and others at the Department of Measurement, Statistics and Computing, Institute of Education, whose interest, patience and considerable support in providing facilities and advice for the author's statistical and programming work, were most helpful in the analyses of the data

(Continued)

Mrs. Daisy Penfold, formerly Senior Lecturer, Institute of Education, whose driving enthusiasm for statistical understanding helped lay the foundations of that understanding in this author and in many other fellow students

The management and staff at the Institute of Education Library, who provided the reading resources of which such liberal use was made in this study, and who also facilitated the computer searches at a later stage in the research

Peter Jobbins, Chief Technician for the Department of Child Development and Educational Psychology, Institute of Education, Anna Brett and her secretarial colleagues in the same Department, and Tom White and his colleagues at the Porters' Lodge, all of whom in their various ways did much to ease the work required for the study

The Advisory Staff at the University of London Computer Centre, for their help with programming problems

Charlie Owen, of the Thomas Coram Research Unit, Institute of Education, whose tremendous programming skills and generous advice helped in the difficult period during which mastery of Fortran and other techniques needed to be learned

And finally, of considerable importance was the support given by:

The Social Science Research Council, which funded two of the author's study years, a privilege which was greatly appreciated

The Central Research Fund of the University of London, which provided financial help towards the costs of preparing research materials

The Minor Research Fund of the Institute of Education, which also contributed to the author's equipment costs

There are many others who could be named. If no person is an island, it can also be said that no research worker functions independently of the rest of the research and academic communities. The final product, in the pages which follow, is an amalgam of the ideas and efforts of a great many people, together with those of the author. To all of these persons the author is deeply grateful.

# C O N T E N T S

	<b>Abstract</b>	<b>vi</b>
	<b>Introduction</b>	<b>vii</b>
<b>1.00</b>	<b>Study Outline</b>	<b>1</b>
<b>2.00</b>	<b>Early Education Attainment: a conceptual model</b>	<b>5</b>
2.10	The development of early reading	7
2.20	The development of early mathematics	17
2.30	The influence of ability	22
2.40	The influence of temperament	27
2.50	The influence of parents and home environments	34
2.60	The influence of pre-school environments	46
2.70	The influence of television	63
2.80	An integrated model	66
<b>3.00</b>	<b>Intervening in Disadvantage: home and pre-school</b>	<b>72</b>
3.10	Educational and social disadvantage	74
3.20	The reality of intervention	82
3.21	Head Start: the American dream	83
3.22	Educational Priority Areas: the British compromise	86
3.23	Three landmark studies	88
3.24	Institutional intervention	92
3.25	Parent intervention	101
3.30	The methodology and research perspectives	108
3.40	Evaluation, cost-benefit, dissemination and policy	120
<b>4.00</b>	<b>Intervention Project</b>	<b>134</b>
4.10	Rationale	135
4.20	Sample	141
4.21	Sampling: concepts and limitations	141
4.22	Choice of sample	143
4.30	Method	153
4.31	Child pre-tests and mid-tests	158
4.32	Parent interviews	178
4.33	Parent programmes	184
4.34	Child post-tests	196
<b>5.00</b>	<b>Analysis</b>	<b>198</b>
5.10	Statistical options	203
5.11	Bivariate relationships	213
5.12	The multivariate approach	216
5.14	Reliability and disattenuation	226
5.15	Validity assessment	236
5.16	The significance concept	251
5.17	Age and time concepts	254

5.20	Multiple regression	260
5.21	The limitations of Ordinary Least Squares	267
5.22	V-ridge regression: theory and evidence	272
5.23	Empirical assessment of the V-ridge model	285
5.30	Path analysis	318
6.00	Results	335
6.10	Child tests	337
6.20	Parent interviews	344
6.30	Parent programmes	360
6.40	Data, sub-sample and group characteristics	384
6.41	Broad characteristics of the data	385
6.42	Sub-sample characteristics	397
6.43	Comparison of parent group characteristics	405
6.44	The experimental design and analysis	413
6.50	Data: reliability and validity	419
6.51	Reliability determination	421
6.52	Validity determination	433
6.60	Data: age and time variables	456
6.61	Age relationships	457
6.62	Time relationships	462
6.63	Derivation of new time variables	465
6.70	Correlations and group scores	470
6.71	Transformations of variables	473
6.72	Group scores	474
6.73	Predictive power of meetings attended	476
6.74	Correlational evidence on other variables	481
6.80	Path analyses	483
6.81	Application of regression techniques	484
6.82	Construction of path models	491
6.83	Specific criteria	505
6.84	Path models 1 to 10	512
	1: Total Disadvantaged Sample	516
	2: Advantaged Sample	527
	3: Disadvantaged Working Group	539
	4: Disadvantaged Reading Groups	549
	5 and 6: Disadvantaged Maths Groups	560
	7 and 8: Disadvantaged Girls and Disadvantaged Boys	576
	9 and 10: Disadvantaged Black Children and Disadvantaged White Children	589
6.90	Cost-benefit indicators	605
7.00	Synopsis	614
7.10	Summary of previous chapters	614
7.20	Comparison of variance predictions	659
7.30	Principal findings	665
7.40	The future?	672
	References	674
	Detailed Contents (Sections, Tables, Illustrations, Appendices)	695
	Author Index	704
	Subject Index	708
	Appendices (see microfiche inside back cover OR separate binding)	

## A b s t r a c t

Theoretical models of early educational attainment are explored and the limitations of some current theories are discussed. Many of the more important intervention programmes aimed at helping disadvantaged pre-school children in Britain and the United States are reviewed and alternative interpretations of some of their findings are proposed. A programme of intervention, based on the parents of 159 nursery children at inner city primary schools, has been carried out for the study and is described here in detail. Various methodological issues are examined in relation to the study's goal of improving the reading or mathematical attainment of these children, when assessed near the end of reception class. The nature of the intervention programme, which has taken the form of eight fortnightly meetings for each of 25 groups of parents, is set out in full. New forms of quantitative indices are developed to define the nomological validities of the individual variables and of the multivariate evaluation models. The study also reports the practical development of a new form of non-stochastic ridge regression, named V-ridge in recognition of the American statistician (H.D. Vinod) who created the original algorithm; it is shown here that this method offers regression coefficients which are far more stable across samples, and often more credible, than are the equivalent ordinary least squares coefficients. Ten path analysis models based on the study data yield a wealth of insights. The models show, inter alia: major differences in the contribution of home environments to the educational attainment of advantaged and disadvantaged children; the importance of the nursery school experience for disadvantaged children; the significant though modest contributions of the parent intervention programme to the outcomes of reading competence and mathematical concepts, but not to the outcome of mathematical numeracy; and the stronger integration of cognitive and academic characteristics in advantaged children than in their disadvantaged peers. The major conclusion is that a nursery class programme of parent guidance, structured and focused on the presentation of practical activities enabling parents to foster children's early reading and early mathematical development within the home, is effective and should be a regular feature of nursery education in disadvantaged areas.

## Introduction

Karl Popper (1976) writes of his "espousal of deductivism - the view that theories are hypothetico-deductive systems, and that the method of science is not inductive...." He proposes "a new theory of the method of science,..... the critical method, the method of trial and error: the method of proposing bold hypotheses, and exposing them to the severest criticism, in order to detect where we have erred.

"From the point of view of this methodology, we start our investigation with problems. We always find ourselves in a certain problem situation; and we choose a problem which we hope we may be able to solve. The solution, always tentative, consists in a theory, an hypothesis, a conjecture. The various competing theories are compared and critically discussed, in order to detect their shortcomings; and the always changing, always inconclusive results of the critical discussion constitute what may be called 'the science of the day'.

"Thus there is no induction; we never argue from facts to theories, unless by way of refutation or 'falsification'....."

The approach of the present study is close to that of Popper, whose deductive philosophy may be seen as an underpinning of the whole corpus of intervention research. The parallels between intervention studies and the logic of the above quotation are strong. In the former there is a recognition of a problem - early and continuing school failure. The attempt to solve the problem through intervention is founded on theories of human responsiveness within a more stimulating environment; hypotheses are formulated to test these theories. Competing theories are critically reviewed in the light of the findings from the intervention, making use of quantitative analyses and other forms of logic. Finally the 'always inconclusive' results of the critical discussion are put forward as the current - but by no means the ultimate - science of the day.

Every intervention attempt in the educational and pre-school situation has been overtaken by more sophisticated projects, with the theoretical conclusions from earlier projects being supported, modified or even rejected in the light of later studies. The conclusions drawn from the present study are advanced in a full awareness of Popper's caution; they stand in the absence of any more convincing explanation.

1.00      Study Outline

A combination of theoretical and practical insights from developmental psychology and early education, aided where necessary by the judgements of advanced statistical assessment, can provide a much fuller awareness of the potential for educational intervention among disadvantaged pre-school children than can be offered by models which limit their focus to the application of single theories. In presenting and testing a theory of intervention in the wider context, this study first examines the theories and reported research on a range of alternative contributors to early educational attainment, and then focuses on what is potentially the most influential of these contributors, the parental environment.

The wealth of research which has already been carried out on the early educational development of the pre-school child, and the potential role of the institution or parent in fostering this development, forms the background to the study. Building on this foundation the study explores and evaluates an intervention programme in which groups of parents of nursery class children at five disadvantaged and one advantaged school in a United Kingdom metropolitan environment are given guidance on the fostering of early reading and early mathematical awareness in their children.

Chapter 2 offers a conceptual model of early educational attainment and suggests that the complexity of the contributors to attainment demands a comprehensive rather than a parsimonious interpretive framework. The differing natures of the two outcomes of interest, reading and mathematics, mean inevitably a duality of focus throughout the study. This is not entirely a burden, since it enables many useful comparisons to be made between the predictors of attainment in the two disciplines. A large part of this chapter is taken up with a discussion of the more fundamental characteristics within the child which are thought to contribute to educational achievement, and the characteristics within the environment which help to foster and potentiate the skills of the child. Among the environments which are studied are those of the parent, the home, and the pervading medium of television. The influence of institutional pre-school environments is also examined in depth, historically and conceptually. A number of current assumptions are questioned.

On the basis of this theoretical background, chapter 3 discusses the widen-

ing range of initiatives in the fields of home and institutional intervention on behalf of the pre-school child. The difficult distinction between educational and social disadvantage, and the confusion between the two when attempting to ameliorate the former, are noted here. A major section reviews most of the large and many of the smaller intervention programmes which have been undertaken in the United Kingdom and United States, as well as a few of the programmes reported from the European Continent and elsewhere. The limitations of several of these studies are a matter of concern, not because of the inevitable methodological problems which remain unresolved in every innovative study, but because some of the fundamental conceptual conclusions which are drawn from such studies are open to question and a different interpretation. A number of the methodological aspects of intervention studies are dealt with in this chapter, including the interpretation of action research and the kinds of conceptual bias which influence both design and sampling. The current debate over the nature of evaluation is referred to, in particular the crudity of outcome measures such as the ubiquitous I.Q. criterion, and also the often disregarded questions of cost benefit, dissemination and the policy implications of what has been learned. Finally a summary of the concepts and hypotheses of the planned study is presented.

The experimental intervention undertaken in the study is described fully in chapter 4 and in the extensive appendices, the latter containing copies of most of the material and instruments used in the intervention project with 159 mothers (or fathers) of nursery class children. The issues of sampling and design receive detailed examination in this chapter. The particular urban environment of the sample is described and the limitations of both sample and design are set out, as well as their potential strengths. Further sections of the chapter describe the many tests used to assess the children before, during and after the intervention phase, the interview protocol used with the children's parents prior to the intervention, and the nature of the reading and mathematics programmes given to the 25 groups of parents participating in the experiment.

A chapter of almost equal importance to chapter 4, in the context of the present study, is the wide-ranging statistical discussion and development in chapter 5. Particular emphasis is laid on the need to subject intervention data to complex analytical models rather than to the somewhat simplistic analyses which are often applied to extensive data sets, masking the deeper insights which may be available in the information. The serious limitations of one particular statistical procedure, multiple regression as defined by the ordinary least squares algorithm, are examined in detail; in its place an alternative non-stochastic method of multiple regression, to which the name V-ridge regression has been given, has been developed in the course of this



study. The rationale and theory of the new method are set out here, with extensive comparisons of the sometimes contradictory results obtained by the V-ridge and ordinary least squares methods when applied to the same data sets. In brief, this chapter sets the statistical background for the analyses of results presented in the next chapter.

Chapter 6 reports on the results of the whole study, focusing first on the child tests, the parent interviews and the parent intervention programmes themselves. The characteristics of the sample and the specific groups identified within the sample are then described, conceptually and quantitatively. Issues such as reliability and validity are dealt with in depth and new concepts of nomological validity and time are applied to the data. The most important findings and insights are derived from ten models of path analysis, in which ten sub-groups within the sample are examined separately. These models suggest that there are wide differences in the contributions made to academic performance by the various predictors in each group. As always, the results of what have to be recognised as innovative methods need to be accepted with some caution, pending replication or alternative studies in the same field of enquiry.

Subject to these caveats, the results indicate that the methods developed for the study offer more profound insights than are usually available from research in this field; they also point to the powerful differences between social class and other groups in the extent to which home environment, nursery schooling, child ability and the child's pre-school academic attainment levels each contribute to the outcomes in reading and mathematics in reception class, 20 months after the initial assessment of children and their home environments.

A summary of the main theoretical and practical findings drawn from the study is presented in annotated form in chapter 7; the implications of the most important of these findings are discussed.

In brief, the study aims to provide some new theoretical insights in the field of early childhood education and to achieve four specific goals:

- i. to establish whether the provision of a short experimental programme of parent education, through group meetings in disadvantaged urban areas, can assist meaningfully in the development of nursery children's early school attainment in the fields of reading and mathematics;
- ii. to formulate and apply a number of innovative methodological concepts in the field of research practice;
- iii. to develop, test and establish the effectiveness of a new statistical

- algorithm for solving multiple regression equations; and
- iv. to develop longitudinal path analysis models which will interpret the differing attainments of children in reception class in terms of a variety of pre-school predictors.

## 2.00      E a r l y      E d u c a t i o n a l      A t t a i n m e n t :                  a   c o n c e p t u a l      m o d e l

---

The many and complex contributors to early educational attainment have long been of interest to educators and psychologists and indeed to all who are concerned about the difficulties faced by those whose attainment is too limited to serve as a foundation for more advanced school education.

Theorists of different persuasions have offered many explanations about the key factors thought to underlie educational attainment. Cognitive, motivational, environmental and other models have been presented in which the factors of interest within a particular discipline are seen and researched as the principal determinants, while other contributors are treated as of limited importance.

This chapter aims to establish the complexity of the contributors to early educational attainment, within the broad limits set by an educational and psychological study.

The first two sections examine the theories on the development of early reading and early mathematics. The postulates within each field differ greatly and this fact is reflected in the discussion of each section. The relative linearity of the acquisition of mathematical skills contrasts considerably with the flexibility of approaches to the acquisition of reading skills. Environment may be a more important contributor to early reading than it is to maths. The section on reading also deals with the sensitive pedagogical issue of whether children should be encouraged or even permitted to learn to read at an early age. A conceptual model on the development of mathematical abilities is presented in the section on that topic.

The next two sections deal with what are often seen as quite distinct features of the child or adult, namely the range of what might be termed abilities or competence, and the range of temperamental or personality characteristics which help determine the extent to which acquired abilities are utilised or further developed. The examination of these two areas is of necessity a brief one, with the main focus being on the contribution of such factors to early academic attainment. An important characteristic related partly to temperament and partly to experience is motivation; this is also examined briefly.

The influence of parents and home environments is reviewed in the following section. The social milieu in which parents in Western society bring up their

children in the early years is dealt with at some length, in view of its relevance to the study presented here. Among the significant issues is the growth of women's employment outside the home and its relationship to the work of child rearing. The cultural norms of early rearing and early educating are noted and some of the research on the parent as teacher is reviewed.

A section of particular importance to the planned study is the influence of pre-school environments. A brief account is presented of the early history of pre-schooling in different countries, prior to a discussion of the problematical issues of what should be the goals of the pre-school experiences provided by society for its children, and whether these experiences should be aimed at providing the children with academic or school-related skills. The focus of this section is on nursery schooling rather than on alternatives such as playgroups, day nursery care or child-minding, although the latter three environments are discussed briefly. Finally some of the wealth of research into nursery schooling is reviewed.

The penultimate section deal with a topic which is becoming increasingly important in the early development of children, namely the medium of television. Although it has been a pervasive influence on children over the past few decades, research into this area is scant. Credible theoretical models have yet to be developed on the influence of television in relation to other influences - rather than television in isolation - and in relation to the differing temperaments and abilities of children.

The final section examines the problem of presenting an integrated research model of early attainment and shows that this task requires both large resources and a wide conceptual framework capable of taking into account different domains of influence. The section concludes by posing a relatively simple theoretical model of early educational attainment in relation to most if not all the major contributors.

## 2.10 The development of early reading

There are continuing debates over the nature of reading, the best methods of teaching it, the age at which reading should be started, and the hypothesised causes of reading failure. Many studies into aspects of reading have focused on only one or two variables rather than recognising the multiplicity of contributors to reading attainment and reading failure. Concern over the possibility of early reading failure has also led to an emphasis by some professionals on the importance of leaving the teaching of reading to the infant teacher, with parents being discouraged from doing anything more than read to their children in the pre-school years. Linked with this approach is the somewhat dogmatic assertion that there are only one or two 'correct' methods of teaching, this assertion in turn reinforcing the advice to parents to leave the teaching of reading entirely to the professionals.

The very large amount of research conducted into the whole area of reading over the past few decades has shown not only how complex are the issues under debate but also the breadth of possibilities for the development of early reading. This section will review briefly some of the theoretical positions on the nature of reading, the debate on methods, the predictive models, and finally the issue of reading readiness and the case that has been made for an early start on reading.

## 2.11 Some reading theories

While reading theorists advance widely differing interpretations of the process of learning to read and reading itself, it is possible to see all these interpretations as adding insight into various aspects of the total process.

Frank Smith (1971,1973,1978) has developed what are now becoming well known theories about some of the main cognitive processes involved in reading. He points out that the prediction of meaning - from the context of the previously read words and the reader's experiential knowledge - occurs in advance of what is taken in by the eye, so that in fact the visual recognition of words merely confirms or disconfirms cognitive expectations, and establishes a further cognitive framework for subsequently read words. In this process redundancy plays a great part, based as it is on existing word knowledge and semantic awareness.

In a discussion of the relation between spoken and written language, Smith (1975) states that conventional assumptions that written language is a visual

representation of speech are wrong, as is the assumption that reading requires the decoding of text to sound; he questions the method of teaching children to blend the sounds of letters, since the prior identification of words or their sounds is neither necessary nor desirable in fluent reading.

According to Smith the fluent reader relies on either ignoring or guessing unfamiliar words, often resisting strong pressure from teachers; the weaker reader yields to this pressure and becomes increasingly confused and anxious, relying less on non-visual information and thus compounding the difficulty. In regard to the learning of reading "the realisation that much of our skill in language lies beneath the surface of the phenomenon suggests that the only way to learn reading is by reading (just as the only way in which a child learns spoken language is by talking and listening)".

A great deal of research has concentrated on the perceptual skills which it has been claimed are a prerequisite for reading. Feiler and Thomas (1980) examine these theories and conclude that the idea that reading is made up of a series of sub-skills, including phonic analysis and word attack techniques, is not born out by research. Skills of auditory and visual discrimination, memory function and sequencing may owe their development in part to reading rather than being prerequisites for it.

One of the most popular theories on reading skills has fostered the idea that training in particular perceptual skills will prepare the child for reading, or alternatively will help a backward reader to overcome what are assumed to be the underlying difficulties in mastering reading. Typical examples of this approach are offered by Mattingly (1972), who writes of inner speech and auditory imagery, Posner et al (1972), who present an atomistic model of the reading process, and Farnham-Diggory (1972), whose exemplar for a kinaesthetic feedback system is supported by what she considers should be a typical example of early reading material: "Get the pot. The pot is hot. Get the pet. The pet is wet. Get the nut. The nut is cut...."

An extensive survey by Hammill (1972) examined 25 intervention projects, half of them based on the Frostig-Horne Developmental Programme of Visual Perception; the author found that the great majority of these studies had concluded that training in perceptuo-motor skills had not led to a concomitant improvement in reading attainment. A later study by Tew (1976) showed that although there was a close correlation between scores on the Frostig Test of Visual Perception and intelligence scores obtained at the same age from a sample of five-year-old children, there was virtually no relationship between Frostig scores at 5 and reading scores at 7.

More sophisticated models of reading, in which perceptual processes are seen to be closely integrated with cognitive processes, are presented by Eleanor

Gibson (1972), who describes reading as a mapping of meaning to the written symbols; Hochberg (1970), who considers the skilled reader to be linking peripheral search processes of the written material with cognitive search guidance; Holmes (1973), who shows that the amount of perceptual processing necessary to identify each letter in a quickly identified phrase is too great for the limits of visual processing and memory, so that in fact context is likely to lead to word identification rather than comprehension being dependent on identification; and Maliphant et al (1974), who offer a major review of the experimental research into the acquisition of skills in reading.

Of most importance for this study is the process by which young children learn to read. Smith (1971) sees the beginning reader as relying initially on a maximum of visual information, but learning gradually how to avoid the memory overload caused by trying to identify successive words in the absence of prediction. Using Smith's insights, one might postulate that a beginning reader only starts reading in any meaningful sense once he or she has achieved the ability to couple information from both the visual configuration and meaning, and that the sudden jump from a laboured exercise on letters and words to reading competence with those same words occurs as soon as a child starts guessing systematically.

## 2.12 The debate on teaching methods

The major debate on teaching methods has centred mainly on the well defined set of highly structured methods in which phonics and word building strategies have pride of place, and on the rather loosely defined methods known as sight or look-say; the latter have much in common with what is termed the psycho-linguistic approach to reading.

There are powerful advocates of the phonics approach and indeed the more formal the educational setting and school ethos, the more likely it is that some form of phonics will be the preferred method. Elkonin (1973), for example, while criticising the analytic-synthetic phonic method as atomistic and mechanistic, considers nevertheless that initial understanding in reading is based on the sound formation of a word, with the smallest inaccuracy in the creation of sound on the basis of the graphic sign rendering the word incomprehensible. He considers that although comprehension is a facilitating or complicating factor in the correct reading of a word, it is the sound language by which its meaning is identified.

Research evidence offers little certainty as to which methods are preferable. A complicating factor may be the ethos and degree of structure in a particular teaching situation. Chall (1967) undertook a major review of the re-

sults of different methods of teaching reading over a period of 55 years. She placed the methods along a continuum from decoding to meaning, and concluded that while differences were not clearcut, they did indicate that programmes with a code emphasis produced better reading and spelling achievements; they also proved better for children with limited mental ability. However she found that the teacher variable counted more in reading achievement than did the specific method used.

Goodacre (1967) examined 100 infant schools and their teaching practices. She compared analytic methods - essentially the meaning approach - with synthetic methods, or the emphasis on phonic strategies. She notes that the initial advantage gained from synthetic methods may not be maintained later when the child moves beyond the level of word recognition. The difficulty with analytic methods however is that they demand a strong home language background, and if this is lacking or very different from school language, progress may be slow. Another study of 27 first grade reading projects (Bond and Dykstra, 1966) concluded that while basal or phonic programmes yielded the best word recognition scores at the end of first grade, differences in regard to comprehension and other skills were less consistent. The authors found that the ability to recognise letters of the alphabet prior to beginning reading instruction was the best single predictor of first grade reading achievement. (A number of subsequent studies have shown, however, that the prior teaching of letter recognition per se does not improve first grade scores.)

A further study (Cane and Smithers, 1971) of 12 infant schools in deprived areas found that successful schools were characterised by the clear organisation of reading instruction from the start, with early phonic teaching; teacher direction rather than teacher permissiveness played an important part in this situation. The Cane and Smithers conclusions emphasise the point made earlier that the ethos and degree of structure are complicating factors, so that it is difficult to determine whether or not<sup>it</sup> is the methods themselves or the teaching environment which account for most of the difference in early results.

The contrasting methods of teaching reading have their parallel in the way that language is viewed in relation to these methods. Downing (1980), whose many years of work on fostering the use of the Initial Teaching Alphabet places him on the periphery of this particular debate, describes the contrasting views on language. He cites Yetta Goodman and Burke as claiming that as a language user the child is not called upon to understand the abstract and complex language and thought processes which are being used in the reading process; Frank Smith too is cited as stating that it is an unfounded assumption that reading instruction should involve teaching children about language.

Downing quotes the opposing view put forward by workers such as Mattingly



and Elkonin. They see reading as a language-based skill in which the speakers' awareness of language as an activity with particular features such as words and phonemes is crucial to the ability to think intelligently about reading. Downing also refers to Lee Cronbach's statement that "the learning of a skill is largely intellectual..... Only after the intellectual work has set the pattern for the task does the excellence of the execution come into its own".

That this debate is far from resolved is seen in the totally contrasting viewpoint of Margaret Spencer (1982), who points out that the achievements of Smith and the Goodmans are only now being appreciated, after they have been swimming for years against the tide of much educational opinion. She sees a move away from systems of instruction towards learning through interaction. The Goodmans, for example, have developed a teaching strategy of miscue analysis, on the basis of which they guide the child's reading development. Spencer also refers to work by Bettelheim and Zelan on the damaging consequences of teacher obsession with letter reversals and other perceptual errors.

While there are a great number of initial reading texts based on essentially phonic methods, there are relatively few standard works which are oriented more towards the development of meaning. One popular and widely used example of the latter is Breakthrough (Mackay et al, 1970), which provides cardboard words and letters that can be used in a 'sentence maker' from the outset. The teacher helps the child to select 'personal' words and encourages her to make up meaningful phrases, then copy them and learn to read them by remembering the phrases; often this is done within the context of a picture which is drawn by the child prior to making up some original statement on the picture. Only as the child progresses does work start on making up words from letters.

### 2.13 Environmental contributors to early reading

Of the various environmental contributors to early reading, the parents and home environment are by far the most important.

The role of the parents in assisting early educational attainment is dealt with more fully in section 2.50. Educationists tend to have rather uncertain views about the extent to which parents can or should contribute to early reading. It is popularly alleged that where parents do succeed in teaching children to read, the method they use is the 'wrong one' because it is not the particular method used at the school to which the child will go. A fault identified by many teachers is the teaching by parents of capital letters rather than lower case ones. The most damaging criticism voiced about parents who teach their pre-school child to read is that their action ensures that the child has nothing

to look forward to in the first year or two at school, and that he will be bored and will become hostile to school. It is difficult for parents, particularly disadvantaged parents, to counter such arguments.

A volume of papers on deschooling, edited by Lister (1974), set out in detail many of these and other criticisms of the nature of present day school education.

What should be noted is that while teacher criticisms of parent initiatives in the field of early reading are often reported by parents at second or third hand, it is seldom that teachers themselves, educational documents or journal articles express such specific views. The more usual approach is that parents are advised to read frequently to their children and give them plenty of 'language experience' and other forms of stimulation, but no advice is offered on how to initiate the child into the art of reading, either before the start of school or during the early school years. Some recent breaks with this tradition will be described in section 2.50.

In the wider community environment the cultural and social background of the child is clearly of major importance and a great many research reports describe the differences in reading performance across social classes or other cultural groupings. The underlying factors are thought to include variations in parental pressure for achievement, in parent, community and teacher expectations, in language structures - the work of Bernstein (1971) and many subsequent commentaries on his hypotheses discuss the language issue in depth - and in the varying quality of the schools attended by advantaged and disadvantaged children.

Among these cultural and social pressures are what Dwyer (1974) has described as sex role standards in relation to both reading and arithmetic achievement. She found that the child's perception of reading or arithmetic skills as sex appropriate or sex inappropriate was a more powerful predictor of sex differences in attainment than were biological sex differences, or even the child's individual sex preferences or its like or dislike of the subjects themselves.

Another aspect of cultural influence is described by Downing (1973) in his comparative review of reading across cultures. An important but as yet little studied contributor to attainment, one which varies considerably between cultures, is the extent to which a culture is dependent on reading. Radio and television have made many cultures less dependent on reading and children having access to these new media tend to regard reading simply as a school activity with little relevance to their own lives.

## 2.14 The prediction of early reading attainment

A great many studies of reading report on the power of a few variables to predict reading attainment over time. Generally such studies focus on a particular group of variables. One variable which does appear in most studies is a measure of social class, as assessed by the status of the father's occupation. This indicator is, of course, only a surrogate for a much more fundamental set of environmental variables representing the parents' educational and cultural milieu and those other aspects of the home environment which have an influence on the child's educational performance.

Three studies reflect the range of wider predictive models.

One of the most advanced is that of Hutchison et al (1979), who present a longitudinal analysis in which reading achievement at 16 years is predicted from a combination of reading scores at 7 and 11 years and from a set of social, family and school environment variables, together with teacher ratings of the children's abilities. The analysis was based on data from the National Child Development Study, a continuing study of a one-week birth cohort in the United Kingdom. The child measures were a limiting feature of this data, for the model had no assessments of the children's reading-related characteristics, such as their meta-cognitive features and their self-perception - characteristics which are known to contribute a fair amount to reading attainment.

Two other studies, those of Stevenson et al (1976) and De Hirsch et al (1966), present predictive models at the other extreme in their selection of variables. Both studies concentrate on a wide range of pre-reading skills, meta-cognitive and cognitive skills, but give little attention to any environment variables other than parent education (in both studies) and family history of reading difficulties (in one study). While both studies compare the correlational strengths of the predictors - in relation to subsequent reading attainment in the early grades - they do not attempt to combine the predictors into a single model to enable a judgement of how each predictor functions in the presence of all the competing predictors.

Despite these shortcomings, the De Hirsch study in particular is a considerable achievement for the depth in which it examined a very large number of identifiable language, perceptual, cognitive, meta-cognitive and other pre-reading characteristics of the child, in relation to both fluent and slow readers.

## 2.15 The case..../

## 2.15 The case for early reading

The concept of reading readiness and the various tests purporting to measure it have long been criticised as having only limited predictive validity. Many of the items in the readiness instruments are basic perceptuo-motor skills similar to those which formed part of the Frostig test referred to earlier.

Sanderson (1963) reviews the attempts to establish the criterion of reading readiness, in part as a counter to the claimed danger of premature attempts to teach reading, and in part as a response to research showing different rates of maturation in children. The author cites Joyce Morris's finding, in her survey of 60 primary schools in Kent, that three-quarters of the heads believed in this readiness concept but in fact relied only on their 'instinct' as to when a child was sufficiently motivated to want to learn to read.

According to Sanderson, since last century there has been an educational emphasis on the 'harm' which can be done to children by encouraging them to read too early and thereby risking failure. The only evidence available in this area is that many young children have not been permitted to learn to read, even though they have shown a desire to do so. No real evidence has been offered on the claimed 'relentless pressure' exerted by parents on their children.

One test which claims a high predictive power is that of Thackray (1965). He offers a composite reading readiness profile, including a set of perceptuo-motor items and some major predictors such as mental ability tests, home environment measures, a picture language vocabulary test, teacher ratings of language, the socio-economic rating of the home, and teacher ratings of the child's emotional and social behaviours.

The history of attempts to block young children from learning to read, aided by the use of readiness criteria of different kinds, is set out in a study by Durkin (1976). Robert Owen in 1816 argued that young children "should not be annoyed with books". Subsequently, Stanley Hall's emphasis on the gospel of heredity emphasised maturity rather than environment as the key to development. This trend was reinforced by the growth of the testing movement, culminating finally in the concept that a mental age of  $6\frac{1}{2}$  was necessary before reading could be tackled. A spurious reliance on correlations as evidence of causation supported this trend, although other researchers were finding that it was mainly the method used for teaching reading which determined at what age a child could be introduced to the art.

Reading readiness tests, according to Durkin, became a popular bandwagon, backed up by reading readiness work programmes. Children were pushed through

these at a speed dependent on their level of readiness as determined by the tests. In the post-Sputnik era there has been more willingness to introduce reading into kindergarten curricula (typically at the age of 5 in American schools), but there has been very little change in methods.

For an acceptable definition of readiness, Durkin cites Ansobel's definition that it is "the adequacy of existing capacity in relation to the demands of a given task", the task in this case being the particular reading method used.

While reading readiness concepts have long served as a barrier against encouraging pre-school children to acquire reading skills, there are also powerful advocates for a lowering of the age at which children are introduced to the art. Durkin is of course an important American contributor to the debate. In England some years earlier Goodacre (1967) pointed out that the stage at which systematic reading instruction could start could be lowered if the reading task was simplified. Even earlier Downing (1963) challenged Schonell's claimed findings that children of 5 to 6, with mental ages of 4 and 5, had been doomed to failure by a too early start on the more formal aspects of reading; in contrast Gates had shown that it was the learning situation and teaching techniques which were the determining factors, so that reading could well start at  $4\frac{1}{2}$  or even 4 in some cases.

The Bullock Report (1975) deals in some depth with the question of early reading. While it cautions against parent pressure and rote-like drill in parent attempts to teach children reading before school, it recognises that there is no virtue in denying a child access to early reading, provided this carries meaning and satisfaction for him or her. If the child is put into stimulating situations with suitably fascinating materials there is no need to fret about the right mental age to start reading.

Southgate (1972) examines both sides of the debate on early reading. As a former teacher she offers a detailed rebuttal of every argument put forward against early reading. These arguments are based partly on the claimed grounds of mental and perceptual incapacity, linguistic inadequacy, the difficulty of the task and the effects of failure, and partly on the belief that even if young children can learn to read, they should not - for example, because their time is better spent on more important activities such as play, creativity and exploration.

The author cites a number of research studies showing success in teaching very young children to achieve some level of reading. Two of these studies found that slow learners and even brain-injured children can be taught to read at ages from three upwards. The author points out that to spend a short time each day enjoying the game of reading does not detract from other activities. In general, early reading and writing gives children an additional mode of both

expression and communication.

The case for early reading is also dealt with extensively by Stevens and Orem (1968). They see no reason why written language should not be learned at much the same time as spoken language, in view of the child's tremendous capacity for assimilating language in the earliest years. Given the use of large lettering, parents can do much to teach the child rudimentary reading in these years. The authors also refer to Durkin's finding that the lower the intelligence level the greater the advantage of an early start with reading.

As Stevens and Orem point out, the large classes and other teaching problems in many infant schools are a further reason for suggesting that parents have an important role in starting their children on reading in the preschool years.

## 2.20     The development of early mathematics

The psychological processes underlying the early attainment of mathematical awareness and mathematical skills are very different from those leading to early reading, although many of the environmental contributors are the same.

There is for example much evidence that whereas reading and language are to some degree located in the left hemisphere of the brain, mathematical operations are located mainly in the right hemisphere. Harris (1975) and Witelson (1975) present important reviews of this issue. It is not surprising therefore that the acquisition of mathematical competence follows a somewhat different pattern from that of reading.

In recent years a fair number of studies have examined the means by which early mathematical awareness and skills are developed in the pre-school and first school years. It is becoming clear that there is a much stronger sequentiality in the experience and learning which underly mathematical attainment than is the case with reading attainment, where a number of different pathways can lead to some reasonable degree of literacy.

There is also possibly a greater maturational element in the ability to handle mathematical concepts than is found in learning to read words and sentences, although the degree to which mathematical attainment is dependent on maturation is itself a disputed issue.

The difficulty in specifying the prerequisites for mathematical attainment is that it is possible to see such attainment variously as a complex of computational skills (whether learned by rote or otherwise), a knowledge of the laws of mathematics, and an understanding of the processes involved in manipulating and interpreting the logic, numbers and symbols which are the defining and operating language of mathematics.

Two interesting studies describe practical approaches to the development of early mathematical attainment. They are worth studying briefly for the insights they offer as to how mathematical skills might be fostered.

Green and Laxon (1977) present an approach to entering the world of number through a series of defined learning opportunities - which they term 'lops' - making use of objects in the home environment. These latter objects are described as 'playthinks' because this brings out their two main attributes, in which the child has to think about what she is doing while playing with the objects. All the 'lops' described by the authors have been devised as a result of observing and working with children, and each one has a purpose which can be explained

in relation to the other 'lops' and the system as a whole. Green and Laxon consider that the development of logical and mathematical reasoning, through these and other learning opportunities, results finally in the achievement of conservation between the ages of five and eight.

Clarke and Mason (1969), on behalf of the Birmingham Educational Priority Area's Action Research Project, devised a series of 67 mini teaching units "to meet the demands from the Educational Priority Areas for a programme to accelerate concept formation". The programme was carried out in three settings in Birmingham (Lines and Widlake, 1971) over a single year, using an eight item test of the conservation of number; it was shown that both part-time and full-time Nursery class children achieved scores on this test which differed at a level of high significance from those of Nursery control children. There was no difference in scores between the programme and control groups in Playgroup settings.

It is noteworthy that both these programmes focus almost exclusively on the attainment of conservation of number as their goal, rather than on the broader aim of mathematical competence.

Nevertheless the skills taught in both programmes do cover a large range of the activities normally found in an early mathematics curriculum, such as learning about invariance, relative and relational terms, and carrying out sorting, matching, one-to-one correspondence, ordering, adding to and taking away from collections or continuous substances, sharing, multiple classification, transformations and other operations on the physical environment.

At a more theoretical level Lovell (1975), who has done a great deal of work on the development of mathematical thinking in young children, brings together a number of papers on this issue, particularly in relation to the acquisition of conservation and measurement. He considers that there is a need for further in depth studies of the thinking processes of advantaged and disadvantaged pupils, with a view to piecing together the nature of the mathematical ideas and information processing which occur in young pupils. "We need to establish..... more knowledge as to the stages through which the pupils pass in elaborating the concepts involved in, say, the numeration system or in the properties of the natural number system". Lovell warns that while mathematics educators must necessarily believe in the particular programmes they advocate, one cannot know the long-term effects of such programmes until well-designed experiments are undertaken.

The author continues: "Beilin..... argues that since the child can construct a conceptual system out of many materials and techniques.... the basic pattern of organisations is internal..... But (put) simply, we want to know what the long-term effects are of particular environmental inputs on the growth



of logical thought and the grasp of mathematical ideas. And can young pupils assimilate particular inputs, as in the case of directed activities in mathematics, without any ill effects?"

Elsewhere, Lovell (1971a) points out that children look upon mathematics as an instrument with which to explore the world and not as a game with arbitrary rules. Many of the ideas introduced to primary school children are understood only in an intuitive and not in an analytic sense.

Soviet mathematicians have developed the science to a very high degree. They have also contributed to thinking about early development in the same field. Skatkin (1971) writes of the use of set theory concepts as the base for teaching mathematics. Logically the notion of a set precedes the notion of natural number; for example, the essence of counting consists of establishing a one-to-one correspondence between elements of a given set and elements of the initial segment of a natural series of numbers. Steffe and Smock (1975) cite the work of Elkonin and Davydov, who consider that the development of psychological processes underlying the learning of mathematics not only do not precede instruction in mathematics but are only formed in the process of learning. These Soviet workers analyse fundamental mathematical structures and hypothesise that cognitive development of children can be altered by school instruction in mathematics; they also consider that the concept of quantity, with its roots in the ordering structures, should be the starting place for school mathematics.

While Lovell and the two U.K. studies described earlier focus much attention on the link between mathematical attainment and the attainment of conservation, there are some critical views on this approach. Mpiangu and Gentile (1975) undertook an experiment on 116 children from 4 to 6 years, randomly assigned to experimental and control groups. The children were administered pre-tests on conservation of number and on a wide-ranging set of arithmetic items suited to the age of the children. The experimental sample was divided into small groups and given intensive arithmetic training in ten 20-minute sessions spread over a period. The other sample was given ten sessions playing a game of snap. The authors showed that both conservers and non-conservers in the experimental sample gained almost equal amounts in arithmetic scores - and gained considerably - compared to the control sample. In both samples the arithmetic achievements of conservers were slightly higher than those of non-conservers.

In the extensive post-Piagetian literature there are a number of important 'training studies' in which various workers have created situations showing that children's scores on various Piagetian tasks can be raised as a result of emphasising certain relevant features of and operations on the environment in question. This is, however, too large an issue to be dealt with in this study beyond mentioning that the rather rigid views on maturation - in relation to

mathematical and reading readiness - which some early Piagetian writers put forward, have given place to a recognition that it is possible for children to undergo acceleration in their achievement of defined Piagetian attainments, and that there is a wide range of ages at which these attainments may be reached. However the issue remains complicated by the question of definitions, since it is still possible to argue that training in specific task skills (and generalisation to a number of related skills) is not necessarily evidence of achievement of the whole range of skills embraced within a particular Piagetian stage - the structure d'ensemble of which Piaget wrote. Much work has yet to be done on the links between Piagetian developmental levels and the complexity of mathematical tasks which can be tackled at any one level.

The sparsity of research on the learning of mathematics in general is discussed by Landsdown (1978), who offers evidence that retardation in mathematics may have any among a number of hypothesised causes, including emotional difficulty, socio-economic factors, poor teaching, cognitive factors, sex differences and the syndrome defined as dyscalculia.

The first four factors identified by Landsdown can be seen as essentially related to difficulties faced in learning and the learning environment - either the child's own emotional or cognitive limitations, or the poverty of the home or teaching environment. Sex differences cover a wide range of hypotheses: reviews by Buffery and Gray (1972) and Hutt (1972) suggest that differential development of the C.N.S. lateralisation of linguistic and spatial skills underlies many of the differences in linguistic and mathematical abilities between boys and girls; on the other hand environmental studies (Blackstone and Weinreich-Haste, 1980, for example) show that schools tend to discourage girls from pursuing mathematics or science at higher levels and this in turn discourages their interest in the discipline.

The concept of dyscalculia is as difficult to define in aetiological or symptomatic terms as is that of dyslexia, and Landsdown (ibid) refers simply to specific mathematical retardation, offering no judgement on the concept of dyscalculia. Workers such as Kosc (1974) have however proposed theoretical explanations for the syndrome. Clements (1976) provides a further review of the issue. The syndrome, if it exists, may well be related to a failure of some essential central nervous system linkage.

A summary of thinking about the development of early mathematical attainment is set out in figure 1 overleaf. It emphasises the sequential nature of development in that field, compared with the far more flexible range of possibilities for the learning of reading. Whether or not this sequence is dependent on the attainment of specific Piagetian stages is an issue that cannot be examined in any detail in the present study.

# Hypothesised development of mathematical abilities in the young child

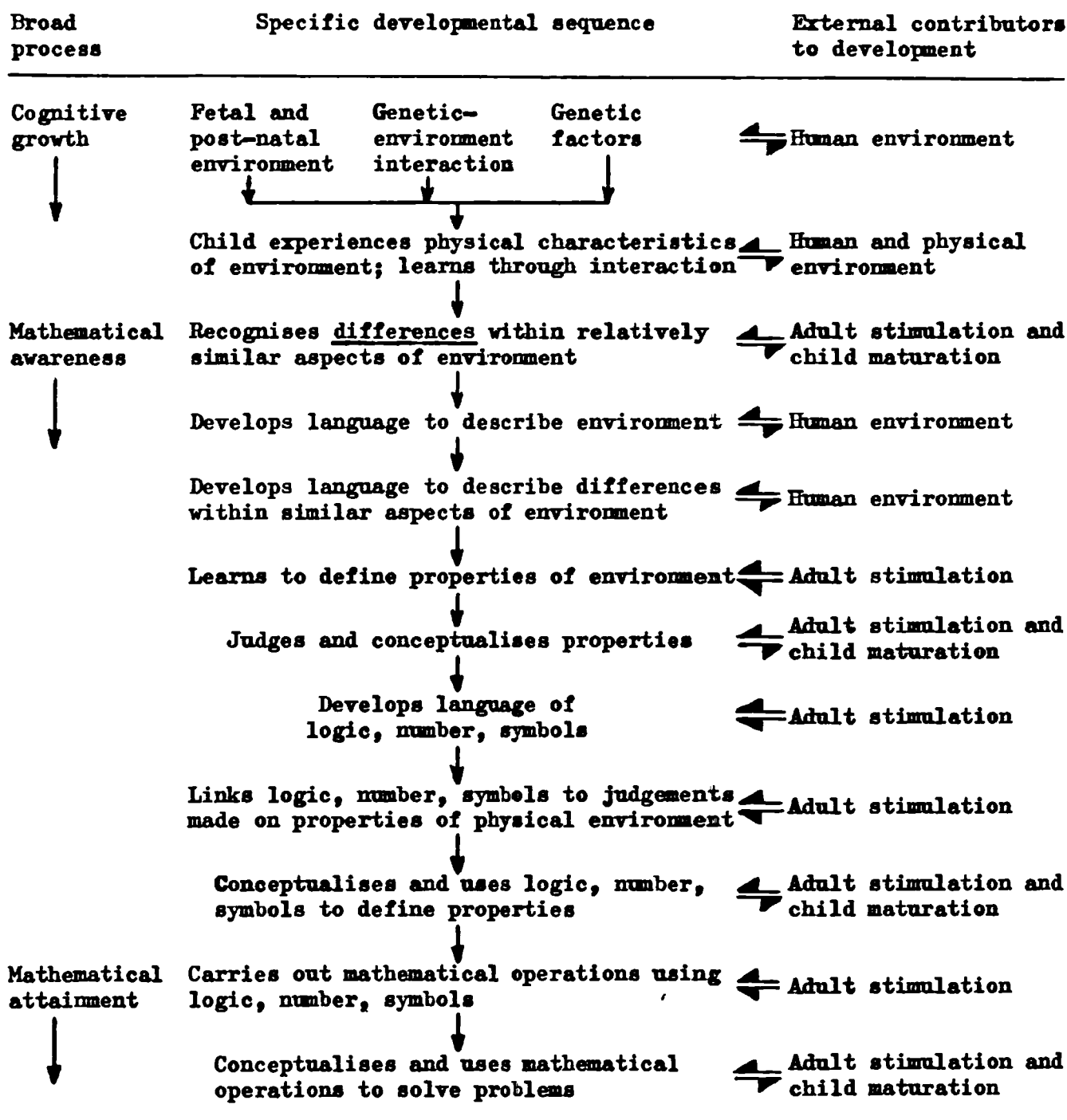


Figure 1. Conceptual model of early mathematical development

## 2.30 The influence of ability

Among the many contributors to early educational attainment the child's ability levels are recognised as crucial and possibly limiting. Whether ability is seen as arising mainly from heredity, from environment and experience, or from an intricate interweaving of all three, it remains an important predictor.

The changing perceptions of intelligence are well illustrated by two books: Butcher (1968) presented a study on the nature and assessment of intelligence in which two out of eleven chapters dealt with the formative influences of heredity and environment; eleven years later Vernon (1979) wrote a study on intelligence in which heredity and environmental considerations were the main focus of interest. The difference lies deeper than the shift of academic emphasis. Butcher's work was a major addition to many earlier studies attempting to identify a relatively stable human characteristic, while Vernon's counterpart was a recognition of the relative fluidity with which this characteristic is expressed as well as of the underlying fact that the characteristic is itself crucially related to the environment which helps to form it and through which its genetic origins are expressed.

A wider interpretation of ability as competence is offered by White and Watts (1973) and White et al (1979). The data derived from an extended study of children's development - the Harvard Preschool Project - were used to formulate a series of social, linguistic and intellectual criteria which in combination served to delineate the behaviour of a competent child at any particular age.

In this study the terms ability, cognitive functioning and competence will be used as far as possible in preference to the term intelligence, to emphasise the wider meaning given to the group of characteristics discussed here.

## 2.31 Influences on ability

If the influence of ability on early attainment is of interest in this study, it is worth glancing briefly at some of the factors which influence ability itself in the early years.

The importance of nutrition, both in pregnancy and in the first years of life, is now commonly accepted. Dobbing (1980) has shown how brain growth in the form of cell growth and myelination of neural connections continues not only through pregnancy as was formerly thought, but well into the second year of life. Tizard, J. (1974) has reviewed the evidence on the close links between malnutrition,

growth and mental development. Although it is considered that most children in the West grow up without suffering serious malnutrition, the work of Cravioto and Delicardie (1972), Cravioto et al (1974) and later researchers has shown that the most damaging effects of mild nutritional inadequacy are caused not through delayed physical growth but through the induced child apathy which in turn leads to a failure to respond to and learn from environmental stimulation. Caldwell (1974) writes of the malnourishing environment as a combination of poor nutrition and an unstimulating social environment; her inventory of home stimulation (Caldwell et al, 1966 et seq.) is a particularly useful measure of various parent behaviours related to children's development.

At a different conceptual level the formidable edifice erected by Piaget (with Piaget and Inhelder, 1969, being only one example of his considerable writings) on the cognitive growth of the child has provided a much deeper understanding of the early years of development and a recognition that this process is a combination of both continuous growth and Markovian jumps into new dimensions of functioning, as the child reaches certain stages of mental operations.

Piaget is not without his critics. Bruner et al (1967) put forward a socio-dynamic model of development in which culture and cognition are closely interwoven and the maturational aspects of development are less clearly delineated than in the Piagetian model. For example, Bruner and his colleagues identified big differences in the achievement of conservation between a group of Wolof children without schooling, and a comparable group of urban Wolof school-children and Western school-children. The divergence between Piaget and Bruner is still wider in regard to the interpretation of age and environment. For Piaget cultural differences manifest themselves mainly in the age at which particular stages are reached; Bruner considers that major differences in environment may lead to qualitative differences in thinking between cultural groups, differences which are not overcome by maturation.

Perhaps Bruner's most important contributions in this area have been his writings on education (such as in 1966, 1971), with his emphasis on the need for structure in learning - for example, offering the child templates which can be placed on experience. He is critical of the attempt to treat discovery as an end in itself.

The importance of play in the cognitive development of the young child is widely recognised. However most of the emphasis placed on the function of play is focused on its perceptuo-motor, social and hedonistic functions. Hutt (1979), for example, speaks of the value of play in early perceptual and motor development, followed by its usefulness in symbolic play and social interaction. But she and many other writers pay little attention to its specific value in cognitive development. The latter occurs provided the environment is rich in

suitable play or other materials and the caretaker sufficiently sensitive to structure the environment appropriately and stimulate the child's interest in such play.

A different contributor to early cognitive development is the interaction with motivational factors such as persistence and creative exploration. Yarrow et al (1975) point out that past emphasis on purely cognitive development has led many enrichment programmes to concentrate narrowly on this goal, unaware that the interaction between the environment and the child's cognitive-motivational functioning is of much greater importance than cognitive functioning on its own. Their research suggested that early environment can encourage a child's motivation to interact actively with people and to explore objects; this may set in motion a sequence of interactions which may be self-reinforcing.

### 2.32 Damaged and delayed development

While a great part of the literature in psychology and education examines the nature of development and its relationship to educational attainment, a smaller but more essential focus is on the situations in which developmental damage and delay occur, leading in turn to slowed or retarded attainment.

Clarke and Clarke (1976) present a powerful case against the assumption that early childhood experience is all important for development. They examine the evidence in detail, referring to a variety of studies which have reported delayed or damaged development. Their conclusions are that early learning is far from being irreversible and that later reinforcements can effect major if not total recovery. Their concluding summary is interesting.

They reject the critical period hypothesis, or even the definition of certain happenings as critical events, and write: "It appears that there is virtually no psycho-social adversity to which some children have not been subjected, yet later recovered"; the quoted phrase appears in italics. It is followed by the unitalicised words: "granted a radical change of circumstances". This latter phrase is of fundamental importance and was seldom noted by the many commentators who quoted the Clarkes' conclusions.

The core of the Clarkes' counter-hypothesis has to rest on the question of how radical a change of circumstances is needed to overcome early damage or delay.

It is known that, given sufficient funds and sufficient research interest, almost any change in the level of child functioning can be brought about over a period. The Heber and Garber (1975) intervention programme was an example of this; the average levels of the intervention children's (low) intelligence was raised by approximately 30 points at a vast cost to the research group and the

application of years of daily conditioning of each child.

What was also not discussed in the Clarkes' challenging study was whether society would ever be prepared to invest large sums in a "radical change of circumstances" for children whose early lives have been spent in poverty with limited or inappropriate stimulation of their development. The cost of later recovery may well be marked on an asymptotic upward curve, with the least cost and the greatest chances of recovery if experience can be put right in the earliest years.

Deutsch (1964) and Hunt (1971) offer more down to earth but perhaps slightly over-simplified prescriptions. While chapter 3 will deal with intervention programmes in detail, it is worth noting these authors' comments on early development in situations of disadvantage.

For Deutsch it is the early deprivation which not only robs children of the opportunity of constructing adequate models of the environment - in Bruner's terms - but also leads to a fundamental discontinuity between the cognitive environment as experienced at home and that experienced at school. The failure of many children to learn in this situation is also a failure of the schools to relate the curriculum to the environmental background and individual abilities and disabilities of such children. Deutsch criticises those sections of the early childhood education movement which have overemphasised protection of the child, restricting the early school stimulation and applying what she terms "suburban mom-ism" to the disadvantaged youngsters.

Hunt (ibid) in turn suggests that there is a mutual interrelationship between lack of competence and poverty - competence seen not as innate potential, whatever that might amount to, but as those abilities, motivations and standards of conduct required for a reasonable degree of academic success. He quotes a considerable number of research studies to support his hypotheses. Factors such as the impoverishment of verbal communication at home may underlie the later deficits in language and number skills; insufficient feedback or incorrectly sequenced reinforcement during interactions in infancy may lead to a failure of persistence in problem-solving at school; and the lack of regular temporal and spatial organisation in daily living may in turn result in an inadequate level of temporal awareness and future orientation. Hunt emphasises however that there are many 'children of poverty' who for various reasons do not suffer in the ways described.

### 2.33 The link between ability and attainment

Most studies which examine the contributors to reading and mathematics attainment find a moderate relationship with cognitive ability, usually as measured by intelligence tests.

The difficulty of presenting an integrating theory linking reading and intelligence has been set out by Swan (1975). The initial difficulty arises with attempting to define these concepts. However if they can be defined by test scores there is clear evidence firstly that the relationship between the two is relatively low in the first years of school, but increases over the years. Schonell's finding that the latter does not occur has not been replicated in subsequent studies, according to Swan. Furthermore, when intelligence is measured in Cattell's terms of  $G_{\text{fluid}}$  and  $G_{\text{crystallised}}$ , it has been found that reading correlates most highly with the latter.

An unusually interesting but still debatable interpretation of the links between intelligence and reading has been put forward by Yule et al (1974). They suggest that although there are clear correlational and regression links between the two measures, it is obvious that there are some children whose reading scores are well above what could be predicted from their intelligence scores, and others whose reading scores are well below prediction. In a study of five survey samples they find that the number of children whose reading scores are below this prediction is significantly higher than could be expected in terms of the normal distribution, with the discrepancy increasing the further the deviance from the middle of the distributions. They consider that this discrepancy may have its origins in children faced with specific reading disability, this being something beyond the normal range of slowed reading attainment.

While these and other studies emphasise the links between cognitive ability and educational attainment, it is more difficult to posit the exact mechanisms by which higher ability is likely to lead to higher attainment in the early years. One can only postulate that within the integrated totality of a particular level of reading or mathematical skills there is a combination of cognitive functions, perceptual skills and motivational determinants within which higher ability may weigh heavily in bringing about a more rapid and less problematical growth of particular educational skills; this in turn is likely to have an added motivational effect in encouraging the learning of additional educational skills, to bear out the finding noted by Swan that the relationship between cognitive ability and reading increases over the years. The same increase over time may well apply to the relationship between cognitive and mathematical ability.



#### 2.40 The influence of temperament

There are a considerable number of behavioural characteristics which describe an individual's unique approach to a particular task or problem. These are in part meta-cognitive variables, having a fairly close relationship with cognitive ability, and in part the expressions of personality. The definition and measurement of such characteristics is a matter of continuing research and speculation and there is little agreement on their nature beyond the common usage of terms such as attention, anxiety, motivation and self-concept.

The question of whether these characteristics are expressions of innate and relatively consistent personality features, or whether they are highly modifiable behaviour patterns linked to the environment to which they are a response, is an issue of some debate among workers such as Mischel (1971, 1973) and Eysenck and Eysenck (1969). While Hans Eysenck does not reject the theory that personality dimensions are expressed as an interaction between innate characteristics and the environment, he can also argue (1964, for example) that there are inherited personality features which can predispose a person to crime. In contrast Mischel rejects any trait theory - such as put forward by Cattell and others - and considers that personality changes over time and situations, although against a background of stable cognitive constructs. Thus personality factors linked to cognitively-oriented personality dispositions are relatively stable, whereas non-cognitive dispositions show high behavioural specificity.

This debate will not be discussed further here, beyond recognising on the one hand that there is a broad though blurred division between behavioural characteristics which are cognitively oriented and those which are a less easily definable expression of personality; and on the other hand acknowledging with Mischel that the expression of these characteristics is closely linked to experience and environment. What is more certain is that the characteristics are of some importance in contributing to the level of attainment achieved by a child.

#### 2.41 Early predictors of temperament

Given the nature of the debate over the origin and expression of the behavioural characteristics shown by individuals, it would be difficult to postulate the nature of the long term development of these characteristics within the young child.

A noted study by Thomas, Chess and Birch (1968) found that it was possible immedi-

ately after birth to identify temperamental characteristics in a sample of middle class New York children, and in a long-term follow-up learned that the early characteristics were of considerable importance in determining the later responses of the children and adolescents to environment and experience; but the authors also found that the outcomes for children with 'difficult' temperaments were closely linked to the way in which the parents coped with these problems and exacerbated or eased them.

The effects of the early physical environment on the development of behavioural characteristics should also be noted. Pasamanick and Knobloch (1960) put forward the term 'continuum of reproductive casualty' to describe the range of neurological damage done to a child both prior to birth and during the birth process itself. They found a great many links with later behavioural disorder of different kinds.

Malnutrition has also been identified as a possible contributor to behavioural problems. Richardson, S. et al (1972) showed how a sample of 74 children who had been severely malnourished in infancy showed not only much poorer class-work than their classmates but also got on less well with their peers and were more likely to manifest behaviour and conduct problems. They noted that the siblings of the malnourished children did not differ from the remaining classmates except in showing a greater level of distractibility. The problems identified in the malnourished children were mainly those of passivity and withdrawn behaviour rather than acting out and aggression.

In general it can be stated that the development of later temperamental characteristics is essentially a function of initial characteristics in interaction with experience and environment in the early years - thus little different in reality from the hypothesised dependence of later competence on a sequence of early interactions with experience and environment.

#### 2.42 Characteristics of particular interest

In an effort to identify those behavioural characteristics which are powerful predictors of educational attainment a variety of studies have assessed sets of characteristics and related these to educational outcomes. Frequently the mistake is made of measuring both the temperamental characteristics and the academic outcomes at the same time, then building a predictive model and claiming that the meta-cognitive and personality variables are causal contributors to current outcome. Even within a longitudinal model of prediction it is never possible to be certain that there is not some element of causation in which the success (or failure) of previous outcomes may have led to an expression or intensification of the characteristics in question, although there are

reasonable grounds for assuming that long-term predictions over time have a modest validity. (Cross-panel analysis offers a reliable method of achieving certainty in this matter for a particular predictor and a particular hypothesised outcome, but it cannot handle the complexity of interactive multivariate predictive models.)

Athey (1970) reviews some of the difficulties in assessing the validity of factors assumed to influence reading. She recognises, for example, how self-confidence can lead to an enhancement of certain cognitive skills, and this can improve reading performance which in turn can boost self-confidence yet again. However there is a fair amount of research evidence that self-concept, autonomy, the accurate perception of reality, awareness of the dynamics of a situation, sound attitudes towards learning and high achievement motivation do contribute to reading attainment.

One characteristic which is highly evident both in the pre-school and early school years is the child's attention span. While this does differ considerably according to the level of interest in a particular activity, it is possible to identify distractibility during the performance of various tasks. Denney (1974) reports on a study of the extent to which conceptual style, Matching Familiar Figures and a measure of attentional style distinguished between good, average and poor readers in four junior school classes. Only attentional style showed a substantial relationship.

However Durkin (1976) considers that children have a 'collection' of attention spans, based on many factors such as mood, fatigue, relevance, familiarity and previous success (or failure) with the activity. The greatest source of distraction for a child is another child; in many classrooms there is at least one boy or girl who spends a considerable time in interrupting the work of others. New acquisitions such as toys or clothes also serve as a source of distraction, often for as much as a whole day.

The cognitive style of the child, as expressed by impulsivity, visual strategies, field dependence and related concepts, incorporates more sophisticated or detailed aspects of attentional behaviour. Kagan (1971) and Witkin et al (1962) have presented a number of studies in this area. However the custom of presenting these or other meta-cognitive behaviours as dichotomous rather than as continuous variables has led to a number of difficulties, both statistical and conceptual, in assessing their importance.

The self-concept of the child is another area of considerable interest to research. While a model such as that put forward by Lawrence (1970) is rather over-simplified - and indeed a later model (Lawrence, 1981) serves to correct the imbalance - the original study did bring to attention the powerful part played by self-concept and counselling in children's success or failure at

reading. An earlier longitudinal study by Clifford and Wattenberg (1962) assessed 185 children's self-concepts and ego strength on entry into kindergarten. Two years later the children's self-concept and ego strength were again evaluated, together with their reading progress, socio-economic background, sex and type of school. It was found that self-concept was a predictor of reading achievement, while progress in reading was not shown to have any marked effect on the formation of self-concept at the post-test level. Although the authors suggested that self-concept might add to the predictive power of cognitive ability measures, the latter were not used in their study.

A conceptual area closely related to this is that described as learned helplessness. Seligman (1975), Dweck and Repucci (1973) and Dweck (1973) describe the serious effects on learning and attainment when children develop a feeling of learned helplessness as a result of previous failure or other traumatic experiences. Dweck and Repucci point out that in children learned helplessness can manifest itself in a belief that failure is due to their own lack of ability rather than to lack of effort.

The above characteristic is in turn linked to the concept of locus of control. Gammage (1974) and Phares (1976) examine this topic in depth and review various measures for assessment of the externality or internality of a child's locus of control.

While the evidence cited earlier in this section points to some of the hypothesised links between birth trauma or malnutrition and later behavioural problems, a far more problematical issue is the question of whether the experience of educational failure is also a contributor to subsequent behavioural difficulty. Clearly there are behavioural problems whose physical or neurological origins are identifiable at an early age. However many maladaptive behaviours only become apparent in the immediate pre-school or early school years and their origins are not easily identified, particularly as siblings from the same environment, for example, frequently do not show the same problems.

Rutter and Yule (1970) tackle this issue in some detail and conclude that severe reading retardation is very frequently associated with anti-social behaviour. Their evidence suggests that both reading difficulties and anti-social behaviour may develop on the basis of similar types of temperamental deviance, but also that delinquency may sometimes arise as a maladaptive response to educational failure.

An early study by Witty (1950) reviews research in the area and concludes that although emotional difficulties are found more frequently in children with reading difficulties, there is evidence that not only can personal therapy help to overcome reading failure but that remedial reading instruction can itself help to reduce the emotional difficulties of readers. Witty goes on to empha-

size that reading difficulties are usually a product of multiple causation rather than of any single emotional problem. While the problems identified by Witty differ in severity from those examined by Rutter and Yule the overall findings are in agreement.

One behavioural characteristic of particular interest is anxiety; educational concern is expressed over this phenomenon, which is thought to be linked in part to the extent to which pressure for achievement is exerted on the children by the school or home environment. Yet it is difficult to assess and a review by Gaudry and Spielberger (1971) shows that the relationship to attainment is somewhat tenuous. It may well be that the manifestations of anxiety are more situation-specific than most other behaviours discussed in this section.

#### 2.43 Motivation

Of all the non-cognitive influences on attainment, few have been given as much attention or enjoyed such concentrated research as has the concept of motivation. Vernon (1969) wrote one of the standard works on motivation as it was then interpreted - following a long period in which atomistic definitions of the characteristic became ever more abstruse, though remaining interesting as theoretical positions. The work and formulae of Hull (1943 et seq.) provide a typical example of the earlier trend. Vernon's own study broke away from this paradigm, emphasising both social and biological determinants; her chapter on motivation in young children referred to the role of curiosity, exploration and play in developing achievement motivation.

Bruner (1969) has done much to apply the theory of intrinsic motivation to school learning. He considers that in view of the high demands made by the school on children it is motivational factors - "the will to learn" - which are of key significance. Intrinsic factors, including curiosity, self-directed attention, competence achievement, interest and learning structure, are of long-term importance.

The importance of the child's role in this learning process has also been stressed by Elkind (1974), but his interpretation differs somewhat. He considers that much schooling tends to make children intellectually burned out and bored; he cites the views of both Piaget and Montessori that intrinsic motivation resides within the child and the child has to choose methods and materials reinforcing to him or herself.

In direct contrast, workers such as Becker and Carnine (1978) have hypothesised that the self-directed learning implicit in programmes founded on Piagetian theories does not usually occur in disadvantaged children. Major U.S. pro-

grammes designed to improve the academic performance of such children have failed because they did not offer structured educational experiences. These and other workers, such as Bereiter and Engelmann (1966), consider that motivation can develop as a result of the child being given activities which yield success as well as being educationally useful.

The fundamental differences between the views represented by Elkind and Becker have important implications for pre-school and early school education, since the hypotheses about how particular groups of children become motivated to learn are influential in determining curriculum policy within any educational system.

Other important aspects of motivation in the context of learning have been put forward by Maslow (1943), who defined categories of children's felt needs, such as the desire for love, esteem and security; by Stott and Sharp (1976), who based their assessment of effectiveness motivation on the child's ability to cope competently with new situations; and by Schultz and Pomerantz (1974), who return to the Hullian paradigm to develop a theory that the competing probabilities of success or failure, as expected by the child, are influential in determining the child's achievement motivation within a learning situation.

Parental attitudes have also been studied by many workers and have been found to be of fundamental importance in the levels and forms of motivation shown by children, particularly in the early years. Vernon (1969) goes into considerable detail on this factor, while Wolf (1966) presents an interesting list of behaviours thought to be indicative of parental pressure for achievement.

For the pre-school child, research by Baumrind and Black (1967) has found that consistent, firm and loving discipline in the home, with high standards of behaviour clearly explained and demanded of the child, is most likely to produce independent, self-controlled, exploratory and competent behaviour in the four-year-old. Although parental warmth and permissiveness are most influential in producing child achievement up to six years of age, after that age parental firmness may be more influential.

For the six to eight-year-olds at elementary school a study by Bühler (1943) found that 80 per cent of children who failed in first grade work had not acquired the ability to work persistently for a goal. Good readers were able to organise their performance, channel their energies and work independently. Bad readers were less mature, lacking in control, wanted immediate attainment of pleasure, and had not learned how to carry out a pre-conceived plan of action. The worst readers were passive, dependent, infantile and showed disorganised behaviour, according to Bühler.

A later study of this same theme is presented by Rosen and D'Andrade (1959), who examined in detail the link between the child's achievement motiva-

tion on the one hand and the family's child-training practices and society's expectations on the other.

While measures such as the Thematic Apperception Test and more objective tests have some value in assessing motivation in the specific context of particular measures, it has proved difficult to assess motivation in a wider framework since one needs to determine this characteristic for each task or group of tasks. The closer the motivational assessment task to the ultimate outcome variable the less valid or independent is the measure of motivation itself. The question of rewards, intrinsic or extrinsic, is also likely to bias the results.

A study by Barker (1976), in which a variety of motivational measures were administered to a small sample of 32 children entering two reception classes, showed that none of the measures used had any unique predictive value for reading attainment by the end of the school year - despite correlations of up to 0.35 over the eight months - once account was taken of the competing predictive strengths of initial attainment, non-verbal ability and several meta-cognitive measures. It is of interest to note that a later reanalysis of the data, using more powerful statistical techniques which had not been available earlier, still showed that none of the purely motivational measures were meaningful predictors, although the relative strengths of the variables which were predictors changed very considerably.

In general it may be argued that while motivation remains an important concept, its assessment is so task-dependent that it is preferable to rely on indicators of related behavioural characteristics such as an overall judgement of educational task-orientation, of self-concept, need for esteem and the external contributors to motivation such as school and parent pressure for achievement.

## 2.50 The influence of parents and home environments

What parents are able to do to influence the early educational attainment of their children is largely a function of the social milieu in which they live and on which are superimposed the child-rearing and child-educating norms of their culture.

This section will start with a necessarily brief review of the milieu in which the young child of today is reared in the United Kingdom.

### 2.51 The social milieu

In recent decades both the social and family structures in the United Kingdom have undergone profound changes. While the social class divisions have shifted moderately, so that the proportion of families with fathers in non-manual occupations has increased, there has also been an increasing level of family breakdown, with the number of single parent families having risen to between 11 and 12 per cent of all families.

Although the prevalence of poverty is largely determined by relative criteria which of necessity alter with changing social and economic patterns and changing consumer demand, the Study Commission on the Family reported (Pond and McNay, 1980) that 1.25 million families are on incomes below the poverty line set by supplementary benefit scales; it is estimated that a further 2.5 million families would also be living below the poverty line if both parents did not work. Black families in Britain are faced with proportionately more single parent situations and higher levels of unemployment and poverty than are white families.

While 0.7 out of 3.7 million mothers of children under five in the United Kingdom were working full or part-time in 1971, five years later 0.9 out of 2.7 million mothers of under-fives were in employment; the proportion of working women in the United Kingdom is the highest in the countries of the European Economic Community. Other figures show that in 1978 there were 14 per cent of children under the age of 1, 25 per cent aged 2 and 3, and 37 per cent of those aged 4 who had working mothers. Half of these children were in some form of nursery or day care, compared with one-third of children whose mothers were not working.

The effects on the child of the mother's employment is a much debated issue. A review in New Society (1981) reports that research has consistently failed to



show any harmful effects on the children and in fact American studies suggest that working mothers are more likely to encourage independence in their children, to compensate for absence by increasing the amount of direct interaction with their children when at home, to have structured rules for their children, and to behave more consistently. Broman et al (1975) have shown in addition that working mothers' children have intelligence scores several points higher than do children of non-working mothers. In the absence of adequate statistical analyses of such data it is uncertain whether these findings are merely a reflection of a bias in the occupational levels of women who succeed in finding jobs (within any one social class), in their level of 'upward mobility' striving, or in their total approach to work-seeking as well as to a better organisation of their children's rearing.

The view that it is preferable for mothers of pre-school children to be at home caring for them in the early years has been put by Leach (1979). She considers that such mothers should not be pressured into going out to work if they do not want to do so; there should rather be State allowances for mothers at home, more flexibility at work for fathers of young children, better community facilities, and more trained child-minders (renamed day-minders or substitute mothers) for those mothers who want to or need to go out to work.

The corollary of the increase in working mothers of young children is found in the pressure for increased nursery and other day care facilities. Play-groups, which provide for some 600,000 out of 1.25 million pre-school children enjoying some form of group, institutional or individual day care, are not seen as the answer since mothers are expected to be present or take some part in the work of the groups. The focus is thus mainly on the alternatives of nursery schools and day nurseries (in combination caring for between 400,000 and 450,000 pre-schoolers) or child-minders (with a total of 70,000 registered places and an unknown but reportedly much larger number of unregistered places).

The adequacy or otherwise of this provision of pre-school facilities has been the subject of a Government-funded report on the services for young children with working mothers, prepared by the Central Policy Review Staff (1978). The report was critical of certain aspects of policy, including the lack of direction, the absence of clear priorities, the fragmented provision and divided responsibility; the resulting situation for parents was described as unjust and inequitable.

A study by Hughes et al (1980) and other research undertaken by the Thomas Coram Research Unit into the provision of nursery care found that a very high percentage of working-class mothers stated, in response to surveys in two inner city areas, that they would like a greater and a varied range of part-time and full-time day care in their neighbourhoods; the mothers said they might well

seek employment if care facilities were available. The majority in this survey rejected the idea of a mother's allowance or mother's wages in lieu of going out to work. The presentation of such a forced choice was perhaps unfortunate, since there is the third alternative of maternal wages as of right, giving the mothers the freedom to choose how to use this.

Even as a political issue there has never been a strong lobby in favour of a State income for mothers as a means of enabling them to go out to work and pay for the day care of their young children, or alternatively to stay at home with the children if they so choose. One author who has advocated this policy is Wynn (1972). In her review of social and fiscal policy in regard to the family she charts the persistent political and governmental neglect of family interests over past decades and examines in detail each aspect of policy and administrative practice. She points to the continuing bias of social security and taxation policies, which have diverted resources away from families with dependent children towards non-parents, and notes that economic forces generally tend to develop adversely to the interests of families if unchecked. Families are not helped as of right and indeed there is a relatively low take-up of discretionary allowances. Wynn argues in favour of paying mothers a wage and says that this in turn would mean that the State would have a greater right to ensure minimal levels of emotional and developmental care for infants. Such a way should be seen as a social and educational investment rather than as social security.

A contrasting means of helping mothers is the payment of child benefit allowances, something which has been introduced into the United Kingdom in recent years. When these allowances are large, as in certain countries on the European mainland, they serve effectively as mothers' wages - especially when there are several children in the family - and serve to give parents the freedom to pay for pre-school care and go out to work or to remain at home if they choose. When the allowances are small, in contrast, they are seen simply as mothers' pocket-money to ease the financial burden of providing for the children and make no real contribution towards the financial liberation of the parents, particularly when employment is not generally available to all women.

## 2.52 Parent-child interaction

The reciprocal nature of parent-child interaction is coming to be recognised as more central to the development of the child than is either the parent and home environment on its own or the individual characteristics of the child herself.

Six authors represent a spectrum of approaches to this issue. Bowlby (1953, 1969) presents an essentially linear model in which interaction is

definable in terms of specific classes of behaviour, such as separation avoidance and exploration by the child and maternal caretaking or behaviour antithetical to caretaking by the mother. He relies largely on the ethological approach to behavioural assessment, leaving little room for the fuller subtlety of organismic interaction. Shortcomings in Bowlby's model, particularly its emphasis on the claimed need of the child for attachment to a single figure, are analysed in some depth by Rutter (1972).

Ainsworth et al (1974) and other Ainsworth studies have focused on the fact that the infant is integrated into the social world from the beginning; the authors see the infant as having a biologically programmed signalling system which activates approach behaviour in adults. They put forward an attachment model which implies a balance between attachment behaviours and reciprocal maternal behaviours, with socialisation as a natural rather than deliberate or pressured process. The limitations of this model are that it ignores individual differences and the conceptual hypotheses of either partner.

Possibly the most sophisticated model of interactive social behaviour is that put forward by Schaffer (1971, 1974), who emphasises that reciprocity is the basis of all interpersonal behaviour. He also points out that a very high proportion of interactive sequences are both initiated and ultimately terminated by the child. Much of the interaction between mother and child occurs through their joint focus on other objects. The value of this model is that it is open to incorporating personality and other organismic variables.

Two models which are more oriented towards the role of communication and play in interaction are those put forward by Richards (1974) and Bruner (1975). Richards shows that social communication for the young infant involves mutuality and reciprocity in the temporal sequence of actions, with faces forming an important source of information and speech-like sounds being understood long before the infant itself produces any speech. What he terms intersubjective communication starts with an exchange of smiles in the first few weeks but slowly becomes increasingly complex in terms of sequences and timing. Bruner (ibid) sees meta-language and play as of overwhelming importance in the early development of the infant's skills and understanding. Like Halliday (1975), Bruner places more emphasis on the function than on the structure of early communication. Elaborative play is another major feature of mother-infant interaction, teaching role shifts and developing an awareness of agent-action-object.

At another level Newson (1974) develops Colwyn Trevarthen's hypotheses on the biological programming of babies' behaviours; fortuitous events are used to develop communication, play and other behaviours in the infant; the way the mother imputes meaning to the baby's responses to these events in turn prompts further behaviours by the infant and yet further ascription of meaning to these behaviours.

While many interactive models are based on limited evidence and are open to various methodological criticisms (see for example Rosenthal, 1973), they have opened a wide panorama of speculative insights into the nature of parent-child interaction in the early years. In general most of these models analyse interaction within the situation of a normally functioning family (though Bowlby and Ainsworth do focus on situations of emotional distress and deprivation).

Rossetti Ferreira (1978) puts forward an unusually interesting examination of the relationship between malnutrition, mother-infant asynchrony and slowed mental development in the child. She points out that in a family in a deprived social environment the child will rarely find in its immediate environment a person available and willing to stimulate and be responsive to its behaviour; at the same time the child itself, because of malnutrition and other ailments, will be biologically less able to stimulate and respond to its caregivers. The combination of these and other factors leads to a cycle of interactional deprivation, which in turns inhibits the child's intellectual development. Thus both the nutritional status of the child and its environmental situation need improvement.

Another unusual study is that of Farran and Haskins (1980), whose detailed observations of interaction in a free-play laboratory situation, using 51 mother-child dyads from two socio-economic groups differing in income, education and occupation, did not support previous work showing low-income mothers to be more controlling than middle-income mothers. While there were considerable differences in the levels and nature of much of the interaction, with low-income mothers spending only half as much time interacting with their three-year-old children, conditional probability analysis showed nearly identical dyadic processes in both groups. A significant finding was that it was the presenting behaviours of the children which prompted the differing patterns of interaction. The authors argued that the differences in the time spent interacting with the children were a function of the value systems within each group. "For a parent to value playing with children as an acceptable adult activity probably requires both the time free from other concerns to engage in such activity and exposure to role models that sanction playing with children as a worthwhile activity."

Other observational studies have also questioned the common finding that parent behaviours differ radically in the laboratory in relation to their social class backgrounds. It has been argued that middle class parents are more likely to be aware of the experimenters' preconceptions and to act up to these, that working class parents are more self-conscious in the artificial environment, and that in any case laboratory behaviours do not necessarily reflect the normal interactional processes occurring in the home. While the last argument may have particular validity, the evidence of the Farran and Haskins study will need replication in other situations before it is possible to reject the weight of

evidence that there are differing behaviours by the parents themselves, linked to cultural and socio-linguistic factors. Streissguth and Bee (1972) present a detailed review and discussion of this question. It can also be noted that studies in which parents under observation were asked to teach their own child and a different child (Bercovici and Feshbach, 1973, for example) showed a generalisation of parent controlling behaviour across children rather than adaptation to a particular child's presenting behaviours.

### 2.53 Child-rearing norms

A major study of child-rearing norms in different cultures was presented by Bronfenbrenner (1974) after he and a team of colleagues had examined parenting and educational practices in six cultures, including those of the Soviet Union, United States and Britain. His study is critical of the lack of parental involvement, companionship and intervention in the lives of their children in the U.S.A. and Britain. He sees Soviet society as far more conformist but also much more caring; there is also a greater cooperation between parents, children and the community, with older children sharing in responsibility for younger siblings or younger school-mates. Upbringing is virtually a national hobby in the Soviet Union.

On the question of intervening in Western society, to alter damaging sub-cultural child-rearing norms, Bronfenbrenner states that research has shown "that any appreciable enduring improvement in the child's development can be effected only through an appreciable, enduring change in the environment and behaviour of the persons intimately associated with the child on a day-to-day basis". Even seriously disadvantaged parents can be motivated and given the skills necessary to serve successfully as desirable models in the development of their children, but they need relief from the burden of sheer survival, the days in search of menial work, the nights keeping rats away from the crib.... for under such conditions they cannot find the time or heart to serve as inspirational models for their children.

Bronfenbrenner is confident that sufficient knowledge is available to prevent retardation and failure. He considers that "ensuring a high level of expertise in persons dealing with the child may not be as critical for furthering the child's psychological development as creating possibilities for those who are potentially the most important influence - parents, friends and associates - to realise their potential". Elsewhere Bronfenbrenner (1975) points out that parent education alone, without involving the children or having them present at the same time, has not been shown to be successful. It is essential that in any intervention visits the mother should be asked to play the primary

role so that her relationship with the child is not undermined. He urges home visiting during the first three years of life, with pre-school education and group care thereafter. He points out that research to date has not indicated that group care in the first few years is effective; on the contrary, results have at times been depressed because mothers come to believe that they no longer play the critical role in furthering the development of their children.

Within the British context a more sympathetic view of child-rearing practices among parents of young children has come from a succession of studies by Newson and Newson (1963, 1968, 1976b), carried out in the same urban community. The Newsons conclude that many widely held beliefs about social class differences in parental attitudes on child-rearing cannot be supported. They use extensive tape-recorded interviews with the mothers of 700 children to examine the reasons for different attitudes and aims and conclude that these are due to home circumstances rather than to a real social class difference in attitudes. However they do identify major social class differences in disciplinary methods and attitudes to literacy, with the latter being a significant determinant of the child's progress through school.

A limitation of the Newsons' major contributions to an understanding of the attitudes and reported behaviours in a wide range of homes is that the authors seldom conduct any statistical analyses on their data; indeed much of the material is anecdotal and unquantified, and although a large mass of anecdotal material pointing in the same direction is fairly convincing, measures of statistical differences on single variables are hardly evidence for anything of importance, given the complexity of environments and behaviours. The attention given to studying attitudes also restricts the possibility of examining actual behaviours in greater depth.

More sophisticated analyses, bearing out some of the Newsons' hypotheses, are offered in studies cited in Schaefer (1972), which show family process to be more highly related to the child's intelligence and achievement than is socio-economic status. Only within the lower socio-economic groups are cultural-pedagogical patterns of child-rearing related to socio-economic differences. However Schaefer also asks whether the results of such studies reflect the parents' response to the child's behaviours rather than the parental influence upon the child's development.

#### 2.54 Child-educating norms

It is conventional - and also legitimate - to label the parent as the child's first and most important teacher. A fundamental difference between the school or pre-school teacher and the parent as teacher is that the latter

is in the powerful position not only of knowing the child very well but also of having the potential to educate the child in a one-to-one situation, even in the presence of a few siblings; in contrast the school and pre-school teacher almost invariably have to do most of their teaching on a group basis - with children whom they can only know moderately well. The difference emphasises the special importance of the parent, even given the likelihood that she will be far less aware of pedagogical and developmental principles than will the teacher.

An early study of the influence of the parents' educational environment within the United Kingdom context was described by Douglas (1964). Parental encouragement as measured by mothers' educational aspirations for their children, parents' interest in and visits to their children's schools, the encouragement given to children to do their school work, and similar factors were important contributors to school performance; children with interested parents pulled ahead whatever the level of their initial starting ability. Douglas criticised the official definitions of social class which ignored the mother's background, since mothers might contribute more to the children's attitudes to learning than fathers did.

The question of the relative contributions of home and school to children's achievement is a difficult one, given the varying criteria of achievement - including factors such as emotional balance, independence, creativity and social awareness. Coleman (1975) offers a reanalysis of the International Education Association (IEA) cross-cultural studies of the effects of home and school variables on achievement. Using a path analysis model on the data from six countries Coleman showed that for 14-year-old children the ratio of direct school effects to total home background effects varied from 0.69 to 0.52, according to the outcome variable; the lowest ratio was that for reading (literature and science were the other criteria). For 10-year-olds the ratio for reading was 0.57. The implication of these findings is that the school effect is little more than half that of the home background in determining the reading attainment of children. The figures for England alone show an even lower ratio of about one-third, emphasising the greater importance of home variables relative to the measured school variables in this country.

While a great many studies examine the contributions to achievement of the home environment, such as parent occupation and educational levels, parent behaviours, the literacy levels in the home, achievement pressure on the child and similar factors, there have been few attempts to provide comprehensive models other than in sociological studies. The latter studies, often based on path analyses, have usually examined the societal and other contributors to secondary school or university achievement. Studies of the contributors to achievement in primary school have tended to focus on a small group of easily assessed predictors, giving a rather unbalanced view of what is a complex model of inter-

related predictors.

In view of the fact that research has identified parental interest in the school and similar measures of parental encouragement as important contributors to attainment, there have been considerable attempts in recent years to foster the involvement of parents in the nursery and infant school.

The Bullock Report (1975) describes the various ways in which parents are now making a contribution to school life, often performing some kind of service for the children or teachers, and often indirectly involved in the learning process. The Report recognises the potential difficulties of this situation but reports on several ventures in which the existence or setting up of outside parent or community groups involving school personnel in the pre-school period has led to greater subsequent parent involvement in the nursery or infant classes.

A wider review of the same issue is offered by Woodhead (1977), who describes the methods of cooperation being tried out in the U.K. and other European countries; these include modification of the organisation and curriculum of the pre-school so as to make it more possible for parents to become involved, with a reappraisal of the roles of parent and teacher in relation to each other and to the child's education.

Tizard et al (1981) offer a sharper study of the same issue and conclude that what teachers expect from parents is not necessarily the same as what parents expect from teachers. The authors report on a research programme to develop parent participation in a number of nursery schools; such participation took place only in terms of what the professionals permitted. The authors were critical of the teachers' role and of their attitudes towards the parents, especially the working class parents.

The problem with these and other studies of proposed or attempted parent-teacher cooperation is that the suggested involvement is essentially one in which parents are being invited to enter teacher territory and contribute in some way to the life of the school - under the guidance of the teacher - or otherwise to be prepared to learn from what the teacher is doing for the child. While it is understandable that a teacher of a class of children should be seen as the focal point around which parents can gather, the fact is that the parent sees herself as the focal point for her children, with school and teacher being but one of a number of external foci for her family.

If the goal is to encourage parents to take a greater interest in their children, particularly working class parents who often have little awareness of how they can contribute to early reading and mathematical awareness in their children, then programmes to draw parents to participate in an environment in which they can never be equals - by the very nature of the expertise required in the school setting - seem inappropriate to the desired goal. Clearly



there is a valued role for the parent who is happy to contribute to the work in the classroom as a lay person with limited teaching or other skills, but that can never be a general model, even for parents who do not go out to work.

Parents may well have a more significant role to play, in relation to school, by being given skills which they can apply in the home setting and thereby co-operate with teachers in the joint task of developing early reading and mathematical skills in their children. This theme is dealt with in the next subsection.

## 2.55 The parent as teacher

Amid the great mass of anecdotal reports on successful programmes where parents have succeeded in improving their children's reading or mathematical skills, usually in cooperation with an enthusiastic teacher or school head, there are some carefully monitored research studies where the value of the parent as teacher has been established. Almost invariably these latter studies concern only the development of reading skills.

Possibly one of the most important studies of recent years is that of Hewison (1981), reported more fully in Hewison and Tizard (1979) and elsewhere. The principal author established, in a study of three samples of working class parents and their children - totalling some 267 youngsters aged 7 and 8 - that after taking account of intelligence and maternal language behaviour, the amount of maternal coaching which the children received was a highly significant contributor to reading test performance, explaining more than a third of the variance in scores.

Based on these findings a further intervention programme was undertaken (Tizard, J., Schofield and Hewison, 1982), involving six schools in a multi-ethnic inner city area. In two randomly chosen schools a parent involvement programme was carried out in one class in each school; in another two schools extra reading help was given with the aid of a supplementary teacher; in the two remaining schools no additional help of any kind was given. Parents in the two parent involvement schools were asked to listen to their children read several times a week from appropriate reading material sent home from the school. This programme continued for two years, from ages 6 to 8. Parents were offered only minimal guidance, except in the case of a few parents who were following potentially unhelpful strategies. At the end of the programme the intervention children were slightly ahead of the national average, while only about a third of the control children were at a level of reading performance appropriate to their age. The performance of children who received extra teaching help was not much better than that of the control group, although the gap had narrowed.

Following this study a variety of inner urban authorities have introduced their own programmes to encourage parents to listen nightly to their children's reading, some schools using report cards which are exchanged each day between the parents and teachers in regard to the progress and problems faced by the children. Initial press reports suggest that these have been a considerable success.

The importance of the new model of parent participation is that not only does it offer a considerable contrast with the normal 'parent involvement' approach discussed in the previous sub-section, where parents are limited to a supportive and somewhat secondary role in the school setting, but it has also been shown to offer a moderately successful improvement in the reading levels of the children involved in the programme.

This suggests that the concept of the school providing all the professional input into the early development of children's reading or mathematics, while the home provides merely an environment in which language enrichment, reading to the child, and useful manipulative activities are stressed, is too restricted an interpretation of the parents' role in the early education of the child.

The potential role of the parent in the early development of the child's reading and mathematical skills is of course much wider than is suggested even in recent studies. Mackay and Simo (1976), McMillan (1973) and Doman (1975) present practical methods by which the mothers of children from two years upwards can start teaching them simple reading skills. Doman claims that his methods have worked even with brain-damaged toddlers. All three books emphasise that the programmes should only be followed if the child enjoys them as games rather than as a serious learning activity.

In one study on early reading, Clark (1976) examined the background of over 1500 7 and 8-year-old children to determine what early factors may have been responsible for the lack of progress of backward readers in this group; she also identified 32 who had been fluent readers when they started school at 5. A detailed examination of this sub-sample showed that while the children themselves tended to have a higher than average verbal intelligence, the parents came from a wide range of educational and social backgrounds; what they had in common was a commitment to education. In general they did not go to any special trouble to teach their children to read; in many cases the children developed their reading skill themselves, on the basis of an initial interest in environmental words. The mothers encouraged the children to look at books and generally introduced them to the local library. Characteristics which distinguished these children from other children of the same parents were their powers of concentration and a social self-sufficiency. Clark says that the experience of these 32 children shows how dangerous it is to assume that the way children are normally

taught to read is in fact the way they learn to read.

What Clark's study does not reveal is how many of the remaining children might have learned to read before 5 had their parents given them the needed experience and encouragement in the last year or two before school. The fact that two per cent of the sample learned to read before 5 without any identifiable teaching effort by the parents is not in itself evidence that the methods used for teaching the remaining 98 per cent are at odds with the way children learn to read. An alternative conclusion might be that the powers of concentration of the talented two per cent enabled them to capitalise on their enriched home environments in the same way that a minority of talented children have taught themselves a variety of other skills.

In summary, a modest amount of research evidence and a growing number of practical initiatives do suggest that parents of pre-school and infant school children have a far more important educational role than is usually ascribed to them. The one-to-one relationship between parent and child in the home is potentially an important situation for learning early academic skills such as reading and mathematical awareness. This in turn leads to the question of whether parents of pre-Nursery and Nursery school children should be given advice on how to teach these skills, or whether they should simply be urged to talk and read to their children and give them a cognitively stimulating environment. The question has important implications for the future direction of early educational policy.

## 2.60 The influence of pre-school environments

The potential influence of non-familial pre-school environments on the development of the young child and the reality of this influence have been issues of educational and social import for many generations. In the earliest proposals for pre-schooling there was an emphasis on the particular value of this provision for disadvantaged children. The modern view of pre-school provision is that it has an inherent value for all children, especially in the last one or two years before formal schooling.

The specific applications of home and pre-school strategies to reduce educational and developmental disadvantage will be dealt with in chapter 3. This section will review briefly the history and provision of pre-schooling in its different forms, and will then discuss the key issue of what are thought to be the role or functions of pre-schooling, concluding with a brief review of some research into this area.

## 2.61 The history and provision of pre-schooling

Since earliest times it has been customary for people other than parents to take a hand in the rearing of the young child. More usually these others have been close relatives of the parents. The better off members of society have used servants or paid attendants to carry out the child-minding function and relieve the privileged parents of some of the burden of full-time caring. An extreme example of this practice were the 'wet nurses' who were paid to breast feed the infants of the wealthy. During the 18th century Western society started providing institutional child-minding facilities in an endeavour to give the children of the poorest segment of the population an early beginning in preparation for the challenge of formal school. This process was accelerated by the spread of compulsory schooling throughout most of Western society in the 19th century.

Two authors in particular provide important historical reviews of the development of pre-school provision. Both describe the ebb and flow of policies and practices in different countries.

Anstin (1976) shows that although created for the benefit of the poor, early childhood education programmes have gradually been adopted by the middle class in at least six of the eight countries reviewed, with provision for the poor being diminished as a result; only in France and Belgium is provision almost universal. This change in emphasis has also meant that the early orienta-

tion of these programmes has changed from a concern for welfare, health and education to a concern for social and emotional adjustment and creative expression based on the Froebelian idea of play. Only in recent years has interest in early childhood education for the poor again gained importance, with special State funds being used for this purpose in most of the countries listed in this review.

Deasey (1978) cites the children's centres set up by Oberlin in Alsace in 1780 as one of the earliest examples of pre-school education. Other early ventures were baby-minding centres set up by private philanthropy in England and France.

The main pre-school development in France took the form of salles d'asiles whose emphasis was both custodial and educational, with strong discipline and rote learning. In a vigorous reaction against the cruelty of some of these centres, educational reforms in the 1870's led to the development of the écoles maternelles - which still exist today. Pauline Kergomard was the chief architect of their philosophy and took steps to prevent them from becoming incorporated in the elementary schools. She stated that she preferred "disorder and noise" in place of the earlier emphasis on discipline and good behaviour, since the former showed at least that the children were alive. There was to be no instruction in these institutions; even Froebelian routines were excluded, although his play concepts were acceptable. Kergomard emphasised that the écoles maternelles were not schools but merely serving to shape the transition from life to school, introducing the child to work and school routine.

A somewhat different tradition developed in Italy. In the 19th century innovators such as the Aggazi sisters stressed training in sense perception and play; other centres emphasised moral training and intellectual stimulation. The most important innovation was that of Maria Montessori who in the early 1900's applied the principles of encouraging independent self-directed activity within a highly structured environment, the individual's freedom being limited only by the collective interests of all the children in the group. Her work was based initially on experience with a group of sub-normal children, but the principles developed there were then transplanted successfully to a centre for seriously deprived children in Rome. The combination of independent functioning and collective responsibility within a structured environment enabled the promotion of a variety of curricula, both developmental and educational.

In Germany the early pre-school centres were focused on strong discipline and custodial caring for the children of working women; in reaction against what was a fairly harsh environment Friederich Froebel set up his first kindergarten in the 1830's, emphasising the importance of play. Austin (ibid) describes Froebel's belief that play is the real work of childhood and the best method by which children learn; it is linked with Rousseau's focus on concrete

experiences and sense impressions. Other aspects of Froebel's philosophy, according to Deasey (ibid), were an emphasis on self-direction and a belief that the child has innate ideas waiting to unfold, with play as the spontaneous expression of thought and feeling.

The United States experimented with each of the major European developments of pre-schooling. In 1860 Elizabeth Peabody founded the first kindergarten, based on Froebelian principles. A rigid adherence to these principles led to a complete break in the early 1900's, with Froebel societies moving towards a totally permissive kindergarten environment. Dewey himself placed much reliance on Froebelian views. With the spread of Montessori's ideas in succeeding decades a powerful movement grew to propagate the latter's more structured principles. Again a rigid adherence to what were seen as Montessori principles led to a major conflict in which two national Montessori movements developed, each with its own schools, the one traditional and the other more liberal in philosophy. A third strand in the growth of the American pre-school environment was that based on Macmillan's Deptford venture, in England, with its emphasis on hygiene and the application of psychological principles.

Anstin (ibid) points out that the history of the early childhood movement in Britain is one of dramatic ups and downs. Robert Owen, a socialist mill owner, set up an infant school in Scotland in 1816, providing for his workers' children from 3 years upwards. He was totally opposed to the mechanical methods applied by some other pioneers and laid down that children were not to be "annoyed with books" but were to be taught the uses and nature of the common objects around them in an environment based on freedom, fresh air and play.

The expanding preschool movement in the second half of the 19th century was based largely on Owen's and Froebel's principles; by 1900 43 per cent of all 3 to 5-year-olds enjoyed pre-school provision. However various reports in the first decade of this century claimed that pre-school was not only not helpful but possibly detrimental to children's development. From 1907 onwards the Board of Education discouraged the enrolment of any children below 5, and numbers dropped rapidly.

There were renewed spurts of enrolment in the two world wars, when it was necessary to employ very large numbers of women in munitions factories, but immediately thereafter facilities were again reduced. Another gradual rise in numbers after the second of these wars was blocked by a Ministry of Education circular (No. 860) in 1960, stopping all further expansion of nursery education and almost ending the training of nursery teachers for the next ten years. The Plowden Report (1967) criticised the government's policy on this issue and urged part-time nursery education for two-thirds of all 3 and 4-year-old children, with provision for 15 per cent in full-time attendance. A Government White

Paper in 1972 gave approval and encouragement for a vast expansion in nursery facilities, though in the past few years financial cutbacks - supported again by claims about the relative ineffectiveness of nursery education - have virtually stopped further expansion, so that by 1980 nursery education provision was still comparatively low, varying from 14 per cent of three and four-year-olds in rural counties to 38 per cent in the most densely populated urban areas.

A significant feature of the pre-school movement in this century has been the development of a sizable private sector. Rachel and Margaret Macmillan set up privately funded nursery schools for deprived children, following the government cutback in the early years of this century. Both women were in the Fabian tradition, arguing that it was unjust to compel children to attend school (at formal school age) suffering delayed development as a result of hunger and ill health. Thus hygiene and nutrition were important elements in their early work, although their later thinking covered a much wider spectrum of educational philosophy. Since then the private nursery school sector has become increasingly oriented towards the needs of more advantaged parents, in areas where free State provision is insufficient or where such parents seek to separate their children from socially disadvantaged children.

Among the other influential English workers in the field of pre-school education in this century were Susan Isaacs and Dorothy Gardner, who stressed spontaneous discovery learning in a blend of the thinking of both Montessori and Froebel. Isaacs considered the role of play to be symbolic and therapeutic, and opposed any prescribed course of instruction for pre-school children.

Thus in England too, as in America, there have been competing pre-school philosophies. Few nursery schools will admit total adherence to a single philosophy although it would be a fair assessment to state that nursery education in this country owes far more to Froebel, Owen and the Macmillans than it does to Montessori.

The strength of the private sector is a highly characteristic feature of English pre-school provision. The largest element in this is the extensive playgroup movement in which mothers and trained playgroup leaders - privately employed by the playgroups - are the only adult influences; more children are involved in these groups than in all the nursery schools combined. While the movement is not exclusively middle-class in orientation, it is more generally the advantaged parents who take the initiative in setting up playgroups. The other major form of private provision is that of child-minding, in which many thousands of registered and unregistered child-minders are paid - usually by mothers who are themselves out at work, or occasionally by a local authority's social services, to look after children in the minders' homes for the whole day. Child-minders take the children of both advantaged and disadvantaged parents,

but tend to offer minimal stimulation. Mayall and Petrie (1981) and Bryant et al (1980) describe the developmental and emotional problems often - though not always - encountered with this form of provision.

While nursery schools and nursery classes, playgroups, other parent-child groups and child-minders provide for the majority of those pre-school children who are cared for outside the home for part or all of each day, a further though more limited form of provision is that of day nursery care, where the emphasis is on basic child care and custody; this latter service is provided for parents who have a social or other special need for the full-time day care of their children, and entry is usually at the discretion of the local authority's social services.

It is important to note that none of these four forms of provision - nursery schooling, playgroups, child-minding and day nursery care - is exclusive.

A mother of a young child may well attend a mother-and-toddler group a few times a week while she is still at home with the child; she may then place the child with a child-minder and go out to work. When the child is old enough the minder may take the child to the nearest nursery class or nursery school for a half-day session, and then fetch the child to continue looking after it for the rest of the day. In other cases children may be left with child-minders until attempts are successful to have the child placed in a day nursery, provided the parents can show pressing social or other needs. In yet other cases non-working mothers may bring their children to playgroups, daily or several times a week, until the children are old enough to be placed in State or private nursery schooling, usually on a half day basis. (There are in addition a small number of nursery centres which now provide both nursery schooling and day nursery care.)

Thus, while the statistics cited in section 2.50 give some indication of the size of the main forms of pre-school provision in this country, they are not necessarily indicative of three or four separate populations of children, each enjoying only that one type of experience prior to formal school. Many children will enjoy more than one form of provision during their pre-school years.

## 2.62 The role and functions of pre-schooling

Given the historical background sketched in the previous section, it is possible now to examine more critically some of the claimed purposes or functions of pre-school provision. In view of the focus of the present study on children in nursery classes, the main emphasis will continue to be placed on nursery education rather than on any of the other three forms of provision.



One of the most important recent influences on the development of nursery education in this country was the Plowden Report (1967) which, as pointed out earlier, led ultimately to a Government decision to start a new phase of nursery expansion. Yet some 15 years later Plowden (1982) announced that she now felt that her previous commitment to the expansion of nursery education had been wrong and that in fact she saw playgroups as the most important avenue of pre-school provision for the future, since they involved parents closely in the running of the groups and were much cheaper than nursery schooling.

The Plowden Report - which dealt with the whole field of primary education in England - had stated that the educational, social, health and welfare grounds for nursery education were strong. The nature of urban living meant that mothers had little relief from their children; alternatively if they went out to work they had difficulty in having their children cared for; both groups would benefit from the provision of nursery education. The education itself could also compensate for social deprivation and special handicaps and could improve language development. However, apart from the single reference to the 'educational' grounds for nursery education and the brief reference to language, all the arguments advanced by the Central Advisory Council for Education (which prepared the report) focused on the social and welfare benefits of this form of provision. The benefits mentioned included physical care, enriched opportunities for play, companionship and new relationships with other adults, and in particular the help this provision would be to the mothers themselves.

Looked at in these terms, the reversal of position on nursery expansion by the chairperson of the Council, Lady Plowden (ibid) is not totally at odds with what was set out in the Report. Many of the benefits described there are also in theory obtainable from well-run playgroups, though they cannot meet the needs of working mothers nor stimulate the children of disadvantaged parents - neither of the latter groups being likely to use playgroups. In her recent statement Plowden points out that playgroups enable far more involvement of parents than do nursery schools or nursery classes; nursery education should serve mainly working mothers and other parents who cannot make use of playgroups. It is for the community to help itself rather than rely further on the State; playgroups serve this purpose.

The value of the playgroup movement and its role in the development of children are issues about which there is little hard evidence. A favourable view of the movement, its achievements and developmental aims is presented in a major study by Crowe (1975), who is national adviser to the Pre-School Play-Groups Association. On the other hand two studies, Rose (1973) and Ferri and Niblett (1977), have examined the progress of disadvantaged children in a number of different playgroups. Rose took 10 playgroups specifically funded to care

for the children of disadvantaged parents and found no conclusive evidence that attendance at the groups reduced retardation. Ferri and Niblett examined the progress of 900 children attending 30 playgroups of different types, visiting each group once a month for five months. Although they found a good relationship with mothers, many mothers did not see the groups as helping social development or learning through play, but simply as useful in relieving them of the children for a period. The authors doubted whether cognitive development of the children was specifically aided. "The conventional playgroup is not one which could offer compensatory provision of this nature." Changes aimed at introducing structure, so as to enable more learning to take place, would be anathema to many in the movement.

On the wider issue of the role of nursery education, some recent perspectives provide a useful overview.

An interesting comparison between nursery education in Britain and that in Romania is offered by Atkinson (1976). In Romania the hours of the mothers' work commitments are matched to the kindergarten provision for their children; there is emphasis on a minimum of training in physical and intellectual skills and a stress on cooperative social development; any child with suspected learning problems gets very high attention. In contrast nursery education in Britain is more open but it also has looser aims, so that goals such as intellectual, socio-emotional, aesthetic and physical development and the effective transition from home to school do not necessarily mean goals suited to the needs of disadvantaged children. There is also vagueness as to whether there should be guided or discovery learning, and a neglect of educational techniques such as structuring of the environment and repetition - both stressed by Piaget, according to Atkinson - due to fear of what teachers see as 'rote learning'.

Another reflection of practitioner thinking on nursery education is provided in a study by Taylor et al (1972) on the attitudes of 578 nursery teachers. Although questionnaires on attitudes do not necessarily reflect behaviours, it is interesting that the teachers give the highest priority to social rather than intellectual objectives, even for disadvantaged children; the intellectual objectives were seen in terms of very general aims such as 'help children reason' and 'take initiative in problem-solving', with more specific intellectual objectives being placed at the bottom of the list of priorities.

Woodhead (1976b) considers that such an ordering of priorities is quite predictable from a knowledge of the traditions underlying nursery school philosophy. Since the beginning of this century there has been consistent opposition to structured or formal preschool education. The recommendation of a 1908 Committee that nursery schools for disadvantaged children should be separate from infant schools is seen by Woodhead as of considerable significance;

it "might be seen as a symbol that the educators of young children were turning their backs on the infant school as a model for their work. The traditional opposition to formal instruction in nursery education has survived to this day... The model adopted by British nursery education may be identified most closely with a belief that the child's development should be allowed to follow a 'natural' course.... central to the European kindergarten movement associated with Froebel".

An example of this approach is offered by Tough (1977) who rejects concepts such as a structured cognitive or structured language programme because they neglect the need for the child to reflect on and use his own inner knowledge and to develop a wide range of strategies in using language for the communication of his own thinking. Tough sees structure more in terms of a linear sequence in the acquisition of skills rather than as structured learning.

Laing (1973), in contrast, considers that the concept of structure is much misunderstood and caricatured. Every nursery teacher structures the classroom in some way; the real question is not whether children need structure or not, but how much or how little is required. It has to be recognised that Bereiter and Engelmann's highly structured methods, which can easily degenerate into mindless chanting, have nevertheless proved more successful than traditional nursery education with disadvantaged children. When workers such as Gray and Klaus have used traditional nursery activities successfully these have been arranged in a logical sequence with a specific relevance for each child, supported by work with the parents and home activities.

Tizard (1976) makes an even more fundamental appraisal of the nursery curriculum. She considers that nursery education today provides much less of the kind of informal teaching which occurred in Robert Owen's nursery centre or in the Macmillans' day centre. The difficulty with the belief that what is taught should be adapted to the needs and sub-culture of the child is that in practice it leaves to the child the determination of educational aims, rather than to society itself, the child's parents or the teachers. This approach may end up denying to the child access to existing skills and values, offering nothing in its place except an easier adjustment to the sub-culture. However it is wrong to see the problem of goals as a difference between free spontaneous play and rigid structure. What is needed is the creative midway of working to a programme or plan of instruction. It has been shown from research that in centres where a specific programme exists, staff spend more time interacting with the children and less time supervising and putting out equipment.

In her examination of thinking in this field Tizard criticises in particular many of the untested assumptions in regard to play. The implication of much educational and psychological theorising and especially the dominant influence

of Froebel within the nursery setting is that children are bound to select those experiences which will best further their development; they will develop skills if they need them, and a teacher should not take a major role in the child's play or creative activities. Yet, as Tizard points out, there is no solid basis of research for most of the common assumptions about play. Current nursery attitudes towards play could well be linked to society's current attitudes towards childhood, where the aim is that children should be 'happy' rather than 'good' or 'useful' as was urged in former times.

While the free play situation is particularly likely to appeal to the relatively rich sectors of society, according to Tizard, other mothers see an almost total mismatch between the values of the nursery school and the working world which their children must one day enter. The strong focus on providing a great variety of play materials and equipment and the high value placed on individuality and creativity means that there is little opportunity for playing with adults or for entering into social rule-bound play with other children, although the ages of 3 and 4 are well suited to that.

In essence the nursery situation is now set up to maximise child-initiated, individualised and free play with objects rather than social play or play initiated or maintained by adults. Yet watching and copying adults has always been a central educational method, with adults talking to the children, explaining, demonstrating and serving as a model. Like adults, children may well prefer a combined work/play situation in which they can play and at the same time be involved in goal directed activities, according to Tizard.

Bronfenbrenner (1979) also offers new insights on early childhood education, in a discussion of the ecology of development. He considers that the capacity of group settings to develop very young children's intellectual and educational competence depends on the extent to which caregivers and staff stimulate and encourage task-oriented behaviours on the part of the child - behaviours such as questioning, instructing, responding, praising and comforting. The ability to foster such development is in turn a function of the appropriateness of the setting, with one-to-one interaction being the most effective for the very young child. The author also warns however that research has shown that children who from an early age are cared for in group settings for most of the day are more likely than other children to engage in egocentric, aggressive and anti-social behaviour both during the preschool years and through later childhood into adolescence.

Bronfenbrenner concludes that the developmental potential of a day care or preschool setting depends on the extent to which supervising adults create opportunities for the involvement of the children in a variety of progressively more complex motor activities and interpersonal structures that are commensurate with

the child's evolving capacities, and which allow her a sufficient balance of power to introduce innovations of her own.

In summing up these and many other views on the nature of nursery education, it is apparent that there is a considerable body of opinion which favours a very loosely structured nursery curriculum, completely distinct from the formal infant school curriculum, with nursery goals defined in rather theoretical terms and focused mainly on general social development, with free choice play as the chief means of fostering cognitive development, and unstructured informal interaction between staff and children as the chief means of fostering language development. In contrast there is a smaller body of opinion which seeks a somewhat stronger degree of structure, with more clearly defined aims and a more specific curriculum focus on language and cognitive development, particularly to meet the needs of disadvantaged children.

Based on the preceding discussion, a description of the role of nursery education or other non-familial preschooling may be given by defining seven principal functions:

- a. Custodial. For many mothers the availability of some form of alternative child care is of great importance, partly to relieve them of the burden of continuous care of those children who do not yet go out to school each day, and partly to free the mothers to go out to work or to go shopping without the worry of having to handle and care for little children at the same time. Society, through its institutions of social administration, is also interested in providing a minimal level of custodial care for those children whose home circumstances are so damaging that it is preferable that the children be separated from their home surroundings for a lengthy period each day.
- b. Socialisation. A great many mothers see the fundamental goal of pre-schooling, especially in its institutional form, as that of socialising their children and teaching them how to associate with their peers and with other adults. Many modern writers on pre-schooling include socialisation in their description of its functions, perhaps because the increasingly small size of the modern family limits the opportunity for natural socialisation within the family. This consideration applies most of all to the middle class family, where the number of children is on average the smallest and where street play with other children is not normally favoured, except in sheltered and safe environments.
- c. Cultural acclimatisation. This function is seen to apply at two levels. For the disadvantaged children in society institutional preschooling is often regarded as a valuable means of acquainting them with the behavioural demands and cultural milieu which underlie the functioning of the middle class school

system into which they will be going. With the increasing level of migration across national boundaries there are also a growing number of ethnic minority groups in many societies. For these children pre-schooling serves a different form of cultural acclimatisation, acquainting them not only with the behavioural demands and cultural milieu but also with the language and customs of the ethnic group which controls the school system.

d. Creative expression. The provision of opportunities for creative expression in one or other form has been seen by most writers of the past and present as a prime function of pre-schooling. Earlier writers spoke of the need for the children to be happy, to enjoy the beauty of nature and in some way to contribute to that beauty by collecting appropriate artefacts; more recent writers see pre-schooling as providing not only the opportunity for happiness but also the materials and encouragement for children to express themselves in drawings, paintings, collages, instrument playing and other forms of creative play, with serendipity playing as much part in the creations as does any deliberate choice by children whose perceptuo-motor and sensory skills are still limited in relation to the chosen tasks.

e. Preparation for formal school. While the general goal of preparing young children for formal schooling underlies much pre-school philosophy, there is also the belief and intention that pre-schooling should accustom children to the sense of ordered and controlled movement which is demanded in most primary schools - for example that there is not an absolute freedom to opt out of or enter any activity as a matter of personal choice, but that for the sake of the effective functioning of the whole group of children there are certain minimal limitations on personal freedom, with adult figures other than ones own parents making requests for performance of certain activities and expecting compliance - preferably out of interest on the part of the child rather than because of any compulsion.

f. Development. The development of children's language competence, cognitive abilities and general knowledge is often thought to be an important function in relation to more disadvantaged pre-school children and indeed this function is often cited today as one of the major goals of pre-schooling. In reality curriculum emphasis is usually laid only on linguistic development, while cognitive and informational goals are an implicit rather than explicit part of the curriculum.

g. Academic preparation. The preparation of pre-school children for early reading or early mathematical performance is often seen merely as an assumed consequence of activities undertaken in pursuance of the other functions described above. Thus, for example, creative expression may enable children to learn

how to hold pens or pencils, while development of language and cognitive abilities and the reading of story books are thought to prepare children for the formal reading and mathematical demands which are to be made on them once they reach infant school. It is much less often that pre-schooling introduces children to reading or to mathematical operations, since there is a considerable body of educational opinion that such activities are premature and may very well prove positively harmful to the subsequent school teaching process. A contrary view is that academically-oriented activities can be a valuable preparation for formal school, especially for disadvantaged children who may not have access to pre-reading or pre-mathematical experiences at home.

### 2.63 Research findings on nursery education

It is impossible to cover in a few pages the many research studies which have been carried out on different aspects of nursery education. Some of the more important studies are summarised here. Many others, focused more specifically on nursery education programmes for the disadvantaged, will be dealt with in chapter 3.

While much research into nursery education has examined administrative issues, teacher and parent attitudes, social behaviours within the nursery, language and play patterns, and variation in the structure and curricula of nurseries, relatively few studies have examined child outcomes in relation to the nature and content of what is provided in particular nurseries. It is only in the field of pre-school compensatory programmes that there is any corpus of research as to the effectiveness of competing curricula. Research on the basic cognitive and pedagogical content of nursery education has also been neglected, compared to the large mass of studies on socialisation, interpersonal relationships, linguistic and other important nursery issues of indirect rather than direct relevance to the school experience which awaits the pre-school child and to the expectations which the school has for children who enter its first classes.

Chazan (1975a) and other workers in a volume of European studies on the evaluation of preschool education describe the instruments which are available for assessing both content and outcomes. The customary problems of testing are exacerbated with young children, according to Chazan, due to the children's limited attention span, their distractibility, their tendency to become bored and their reaction to strange (testing) situations.

A comprehensive review of the state of research into pre-school education in the United Kingdom in the mid-1970's was provided by Tizard (1974). Her report noted only a handful of studies on the effects of nursery schooling; there were more studies examining child and staff behaviours, the factors con-

tributing to those behaviours, and other situational aspects of the nursery environment. Tizard discusses the areas which need further research, such as the potential role of the parent both in the pre-school environment and in the home itself.

The results of a number of United Kingdom studies on the effects of nursery schooling are described by Chazan (1975b). One study which concentrated on broad indices of socialisation found no overall improvement in nursery children, compared to matched non-nursery children, one year after the initial assessments had been made; however communication skills were shown to differ. A similar study using matched groups found, in a post-test assessment of infant school children between six and seven years of age, that children who had been to nursery school had a higher social quotient (on the Vineland Social Maturity Scale), but there were no significant differences on the English Picture Vocabulary Test.

Chazan also reports on the findings of Douglas and Ross (1964) in regard to the progress of a sample of 224 children who had attended nursery schooling at 4 and were subsequently assessed on batteries of educational and cognitive tests at the ages of 8, 11 and 15. This sample, consisting mainly of children who by reason of relative disadvantage had been admitted to pre-school, formed part of the National Survey of Health and Development; their performance on the tests was compared with that of the remaining 5,000 children in the national cohort. Despite the fact that the nursery sample was more disadvantaged, their ability and educational attainment at the age of 8 were higher than those of the average survey child; by the age of 15 however the nursery experience children had fallen slightly behind the rest of the cohort. None of the differences was significant. Douglas and Ross also found that several indices of adolescent maladjustment showed the nursery children to be facing more difficulties, but hypothesised that without the benefit of the nursery experience these difficulties might have been worse.

An observational study by Tizard et al (1980a) examined the language usage of four-year-olds at home and at nursery school, in the light of the recommendations of the Bullock Report (1975) that working-class parents should be urged to 'bathe their children in language', and that nursery and infant teachers should give measured attention to such children's language needs. Transcripts of the recorded conversations of the children and adults (using inconspicuous microphones, with no observers present) showed that the quality and range of language interaction at home were considerably higher than at school. A much deeper exchange of meaning occurred in some of the working class home conversations than in those recorded at nursery school with the same children. On the other hand the social class differences in the language used in the 15 middle class and 15 working class homes were very small or absent, according to the different criteria used. The authors emphasised that in a class of children the teacher could not



be expected to hold extensive conversations with a single child in the way that a mother can.

Another area where assumptions about the effectiveness of nursery education has been questioned is on the issue of play. Gehlbach (1975) notes that much kindergarten or nursery play has no clear rationale and no research justification. There has been no assessment of the learning effects of 'educational play' as provided in the nursery environment, and yet criticism of the role of play generates hostility against the questioner. In every other area of schooling claims are put to the test. There is no evidence to date that play helps creativity, motor skills or social development. Gehlbach suggests specific research on the development of forms of play which involve self-directed learning activities, with accompanying strategies to make play materials more complex once they have been mastered in their simpler forms.

A detailed study of play among more than 100 children in twelve pre-school centres is reported by Tizard et al (1976a, 1976b), again using observational techniques. Play was assessed in terms of the materials used (e.g. partial, appropriate, symbolic), the levels of complexity and the levels of social interaction, ranging from solitary to cooperative. The authors noted that much of the play in the pre-school setting was at a rather low and simple level and activities were seldom pursued for any length of time. Factors which seemed to militate against a longer attention span were the great variety of alternative play materials, the lack of staff pressure on the children to persist, and the distraction offered by large numbers of other children. "Thus, although it is often argued by educationists that the intrinsic motivation of self-initiated play leads to the kind of self-absorption which is the best guarantor of learning, in practice other aspects of the free play situation tend to prevent such absorption." Only rarely were staff involved in the play situations - even with potentially complex games - although they did talk to the children about what they were doing.

The authors also found that the larger the 'open plan' situation, the less the socio-imaginative play. Working class children were more than twice as likely to play outside as were middle class children, and did not make as much use of play materials as did the latter group. On the other hand in language-oriented nursery schools working class children scored much higher on language tests and engaged more often in dramatic symbolic play than did other working class children.

The importance of the above study and other contributions from the research team at the Thomas Coram Institute in London lies in the fact that many of the theories which are central to the thinking of policy-makers in the field of nursery education have seldom been researched or fundamentally challenged before

now. It has to be recognised of course that sensitive research evaluation of the kind undertaken today would have been impossible in earlier decades when the foundations were being established for today's nursery schooling. But the paucity of research in the area is perhaps evidence of how completely the assumptions of some of the early pioneers have been accepted; in contrast the effects of formal schooling and the nature of the curriculum and school environment have been subjected to intense questioning and research.

Other work on the nature of pre-schooling in Britain is that undertaken in the late 1970's by the Oxford Preschool Research Group under the direction of Jerome Bruner. The Group's five studies came to some interesting conclusions. Sylva et al (1980) found that activities with a clear goal structure, such as construction tasks, jigsaw puzzles and free expression art, were associated with complex play and long spells of concentration. Small centres with high staff-child ratios showed more achievement; adult involvement had a positive effect on children's play, raising it to more complicated levels and keeping it going for longer intervals. Another team in the Oxford Group, Wood et al (1980), noted that the adults' style of interacting in various pre-school settings went some way towards determining the apparent competence of the young child. A third team, Bryant et al (1980), found that the quality of child-minding in the large number of child-minding situations examined was disturbing; seven out of every ten minded children seemed not to be thriving, being quiet, detached and passive while at the minders, unlike their behaviour at home.

In general, Bruner (1980) is critical of the decade and a half of executive promises in Britain, echoing the earlier criticism of the Central Policy Review Staff (1978). Bruner too urges a coherent national policy in the pre-school field.

Both the Oxford team's research (Smith T, 1980, and a separate intervention study reported in Tizard et al, 1981), noted low levels of parent involvement in nursery schooling. This was seen to be wasteful of parents' interest, since parents were found to show considerably greater interest in their children's pre-school experience than was thought by the staff in charge of the children's groups. The aim of the Tizard intervention study was to get parents to understand nursery education and to change what they did with their children at home. This had only a limited success - after two years of intervention - although staff-parent relationships became exceptionally warm. One of the difficulties was the belief in professionalism by both teachers and parents, as a result of which both groups felt that education could be left to the teachers.

In a summary comment on a number of European pre-school evaluation studies, including Chazan's contribution on the U.K., Stukát (1975) finds that the overall results from the various studies are not conclusive. The short-term effects have often been positive, but not always so, and in the very few long-term

studies that have been conducted it is only exceptionally that effects have remained for more than a year. Widlake (1975) offers an equally gloomy summary of research findings on the effectiveness or otherwise of nursery education and argues for a radical re-examination of the pre-school curriculum, in place of the sentimentality and "semi-theological basis of some current nursery orthodoxy".

There has been little evidence to challenge Stukát's findings in the years since the European report. The work of Tizard, Bruner and others in recent years adds emphasis to the questioning voices.

There is of course a possible alternative explanation for the negative findings. The evaluation instruments themselves may not be powerful enough, or it may be that the wrong skills are being assessed. While the criterion of cognitive ability - often described as I.Q. - is inappropriate as a measure of the effects of pre-schooling, since it can be assumed that the genetic component in this complex product of genes and environment is an abiding influence, the more obvious criterion of how well the child performs at school in the basic skills of reading, mathematics and social behaviour are rightly seen as outcomes which should be influenced to some degree by the pre-school experience.

The increasing range of test instruments described by Chazan (1975a) and others suggests that there is no real problem in evaluating these skills, even if only approximately with restless young children. Two other fundamental aspects of evaluation are the research design and the statistical techniques used to analyse the data. Again the evidence in the volume of the Council of Europe (1975) and in other studies on effectiveness indicate the reasonable adequacy of most of the designs described in the literature. What are not fully explored, there or in most other experimental studies, are the possibilities for using more sophisticated statistical models which take into account the large number of contributors to educational attainment. Most analyses of research data, especially intervention studies, are based on relatively simple techniques of analysis, using for example t-tests to examine a range of profound differences between samples or relying on problematical techniques such as matching in an attempt to equalise samples. These and other methodological issues will be dealt with at some length in later chapters of this study.

If the validity of most pre-school evaluation can be accepted, the only possible explanation for the repeated findings that nursery schooling has little or no long-term effects on school attainment or other measurable outcomes may have to be sought in the evidence presented here, namely that the prime and expressed focus of much nursery schooling is on play and the creation of a happy environment for the pre-school children. Whether the children obtain more enjoyment from play as a result of this institutional experience has not yet been tested. It does appear however that the nature of the play situation, where

the emphasis is on unstructured activity chosen and pursued by the child with little or no adult guidance, and usually with little adult involvement, does not generalise to improving the child's motivation, learning skills or those other aspects of competence which are thought to contribute to educational attainment in the subsequent formal school setting.

If this is indeed the case then it is understandable why many educationists in the field of nursery schooling contend that the goal of providing pre-school children with a happy environment is and should be a fundamental justification for that schooling. There are however other implications of this contention, such as the level of training and investment that can be justified to yield such a result.

If the policy in nursery education is essentially that of aiming to facilitate play and happiness, with social, linguistic and educational progress seen as inevitable concomitants rather than deliberate goals, there may be little reason for questioning the reversal of position by Plowden in favour of play-groups for the majority, with nursery investment focused only on disadvantaged children - the indirect goals ~~could then~~ become direct ones in order to help overcome these children's developmental and educational limitations.

Such a change in policy would mean a return to the goals of those who founded the nursery education movement - for the sake of the poorer rather than the richer elements in society.

## 2.70     The influence of television

" The box that talks but  
never waits for an answer "

- a Kenyan description of the radio

The fact that young children in Britain and the United States spend between 20 and 35 hours a week watching television, a medium even more thrusting than that of radio, has inevitably drawn a great deal of research on the hypothesised effects of this viewing.

The advertising industry has done a great deal of its own, often unpublished research. It is possible to conclude that the continuing high level of advertising aimed at children is evidence that by and large the commercial message is judged to have been effective.

An early study by Schramm et al (1961) on 6,000 American children has been seen as definitive because it included children who had not previously viewed television; today it would be difficult to find such a sample in any industrialised country. Schramm found that elementary school children with the highest marks in school were also the heaviest viewers; he also concluded that television helped low achievers to develop vocabulary skills. Around the age of 10 the brighter children reached 'saturation point' in their viewing habits. A more recent British study (Darden-Smith, 1978) examined the habits of 1,600 children aged between 7 and 17; it found that nearly half the 7 to 10-year-olds watch up to 9 p.m. The study examined the degree of parental control exercised over the children's viewing and found that in 79 per cent of families there was no control whatever. .

The hypothesised effects of television (and to a lesser extent radio and comics) on the socialisation and social and ethical values of children is an area of particular concern to legislators, the churches and education as a whole.

Anderson (1978) examined the social effects of television in an extensive review of all the available studies carried out between 1956 and 1976. These showed a weak positive relationship between viewing violence on television and subsequent aggressive behaviour.

While Anderson included only those studies which had an adequate design, it is seldom that any studies have examined the hypothesised effects of the media within the framework of a multiplicity of other predictors, including personality.

Relatively few studies have reported on the effects of television on aca-

demic attainment, although a good deal of research has been carried out on the educational and general effectiveness of the long-running American programme Sesame Street. The British response to the ethos and methods of Sesame Street has been mixed - it has been shown regularly on Independent Television over a number of years, but Monica Sims, then head of children's programmes on the B.B.C., expressed hostility because of what she termed the "authoritarian aims" and indoctrination of the programme (The Times, 1971).

Research undertaken on Sesame Street in the U.S.A. tends to show a generally favourable effect. Lesser (1977) offers a wide-ranging review of the effects of all television on the pre-school child. He points out that there is evidence that it is the interaction between the home environment and the Sesame Street programmes which may be more important than the programmes themselves. The causal relationship has not been firmly established, since attempts to 'intervene' by encouraging some children to watch Sesame Street and not encouraging others, have not been much of a success. His own conclusions are more gloomy than those of several of the major studies which he quotes. Among the effects noted for the Sesame Street series were increased achievements in letter, number and object relation skills.

Lesser is generally critical of the effects of television on children's development. Studies he quotes have found that children's comprehension of what they watch is surprisingly low given the number of hours a day spent on viewing. Although the medium appears to have a considerable potential for instruction, it is a one-way communication device, with the children being passive recipients of information. In the area of cognitive development the type of active engagement possible through television is particularly limiting, given the special nature of preschoolers' reasoning abilities, the absence of engagement at the physical level, and the lack of any immediate feedback on the correctness of the children's solutions to problems. There is however more potential for the use of television in language development, if programmes are specifically designed for that purpose.

Another American study (Busch, 1978) examined the effects of television on children's reading performance, using a sample of nearly 600. Busch found that in the early school years parental influence on reading performance was at its highest. She also noted that by the time children reached grades 4, 5 and 6 a pattern began to emerge in which the slower readers became less enamoured of the printed word and came to rely increasingly on television for information. Fast readers succeeded in combining viewing of the more exciting episodes on the screen with returning to their story books during the less interesting periods.

A smaller U.S. study (Perney, 1978) used a sample of 200 kindergarten children from middle-class suburbs, examining the effects of a variety of television programmes. His study found that time spent watching The Electric Company

- a popular educational programme appealing to slightly older children than does Sesame Street - was positively related to school achievement, whereas the time spent watching comedy, variety and drama shows was negatively related to achievement. On the other hand no significant positive correlations were found between early school achievement and the family's interest in the child's television viewing habits.

Ghikas (1978) reports on a British study of the effects of You and Me, an educational series designed for pre-school and infant school children; a comparison between 41 five-year-olds who watched a You and Me series at school over a few months, and 40 children who did not watch the programme, found a highly significant gain for the experimental children on a composite measure reflecting early reading skills, spatial ability, several Piagetian measures and English Picture Vocabulary Test scores. The author points to the difficulty of identifying whether these gains were or were not due to an interaction effect with the teachers and schools.

Overall the conclusions from these and other studies appears to be that while viewing of educationally-oriented programmes does have positive effects on the academic attainment of young children, these effects may be mediated to a considerable degree by the environment in which the viewing occurs, implying thus that the interaction variable may be more powerful than the viewing itself. The whole area of interaction effects between television viewing - the length of viewing and/or the choice of programmes - and the child's own cognitive and temperamental characteristics is another area which has yet to be explored in any depth. It is possible that these latter interaction variables may prove to be more powerful predictors of social and academic effects on the child than are any of the other television-related variables.

2.80     An integrated model

Explanatory models of pre-school and early school attainment are often deficient, since they present only a limited array of the many contributors to attainment. This is due partly to a lack of statistical sophistication, but even more to the narrowness of the conceptual model, with an over-emphasis on one or other theoretical position. Thus an educational model will stress variables such as the verbal and non-verbal intelligence quotients yielded by some test, a previous academic attainment, teacher and other school variables, and perhaps a single measure such as social class representing the parent environment. In contrast, an environmental model of early attainment may include parent occupation and education, ethnic group, quality of housing, parent pressure for achievement, school staffing and ethos, but only a single measure of domains seen to be of less interest, such as the child's ability. Another approach might focus mainly on temperament and personality variables, ignoring most of the other domains. Yet another view could stress a series of birth, health and nutritional variables.

Frequently the number of variables involved, even in these partial models, is so large that recourse is had to relatively crude methods of combining variables and then testing the composite score against the outcome variable while 'controlling for I.Q. and social class', it being assumed that this procedure deals effectively with most of the remaining sources of variance.

There are two important considerations in presenting an integrated model of predictors of a specified outcome. Firstly, while the ideal would be to assemble all conceivable variables of importance in every major domain, this is clearly impossible in any but the most comprehensive research situations. It is however possible to select a few important variables from each domain and use these as representative of the larger groupings, recognising the limitations of this approach. The second consideration is that it is not always easy to obtain values from variable domains outside ones particular field of interest. Thus a health oriented study of development may have to rely on crude indicators of home and school; an educationally oriented study may find it difficult if not impossible to ask for health records stretching back to birth, and equally may not have the entrée necessary to assess the home environments of the children under study.

The task of presenting an integrated model of early educational attainment is thus not an easy one and it is understandable why there have been so few attempts to undertake this work.



Another shortcoming of many studies is that they are based on a dichotomous model of psychological or educational handicap, with the focus on learning disabled or backward children in contrast to a group of 'normal' children. This leads to attempts to identify those variables whose mean levels distinguish between the groups at some specified level of significance. Again it is understandable that, in a situation where the causes of educational failure are the prime concern of the investigator, analyses which maximise the differences between groups across one or a set of variables should be of particular interest.

While such models may serve to highlight particular variables of potential importance to the handicap in question, they ignore the reality that children vary individually across the whole range of predictive variables; grouping children as 'normal' or 'handicapped' may offer some conceptual and statistical shortcuts, but it also serves to cloud the essentially continuous nature of almost every predictor and outcome variable. It is only at the point of extreme breakdown - the collapse of immunity in a biological organism, or a final and total loss of interest in a school child who has a long record of failure - that a Markovian process operates to precipitate an individual into a different category of functioning. Even such a radical change may apply only to selected areas of the child's functioning. In the ordinary school environment, and certainly in the early years when most children still cherish some hope for their success in the school environment, the research dichotomisation of children into failing and succeeding groups is highly problematical.

Among the recent studies which have endeavoured to look at the wider spectrum of contributors to early attainment are the three national cohort studies carried out in the United Kingdom over the past four decades. These studies have followed up to 15,000 children, using one week birth cohorts taken every twelve years since 1946 - gathering data on the children and their parents at birth and at regular intervals in the succeeding years. When the children reached school age additional data were gathered on the school environments and attainments. Douglas (1948, 1964), Davie et al (1972) and Chamberlain et al (1975) offer examples of the large number of reports produced on the work of these surveys. The studies have helped to highlight the importance of medical, social and educational factors in the early attainment of children. Their value lies chiefly in the comprehensiveness of the data bases which have been assembled, and the possibilities which they have opened for presenting macro models of development.

Broman et al (1975) presents an equivalent American study. Although not based on a nationally representative sample, the data on more than 26,000 children born and assessed in twelve leading maternity hospitals in this collaborative study were compared, inter alia, with findings on the children's I.Q.'s at the age of four. A particularly interesting finding here was the importance of

mother's education as a predictor, alongside the more well known pre-natal and post-natal variables.

The International Association for the Evaluation of Educational Achievement (IEA) has also yielded a large amount of data on children's environments and their academic attainments at different ages in a number of collaborating countries. Coleman (1975) and Noonan and Wold (1977) present re-examinations of the IEA analyses, showing that path analyses of the data provide more credible explanations of attainment than do the original models and, contrary to the IEA findings, indicate that the home environments contribute more of the variance in academic attainment than do the school environments.

In the context of the United Kingdom at least two large educational studies have also presented global models of attainment. Peaker (1967b) analysed the data gathered on behalf of the Plowden Report (1967). His use of stepwise regression, combined with path analyses, suggested that the home environment was overwhelmingly more important than the school environment in predicting attainment in junior school. Rutter et al (1979) examined the school environments of a large sample of children at twelve secondary schools in an inner urban area, following the children longitudinally for a number of years. The authors conclude that school process variables are of major importance in determining both academic and behavioural outcomes - the latter including delinquency and other indicators of maladaptive functioning.

These and other studies have offered important insights into the relationships between the considerable number of variables which predict educational or social outcomes. However a particular limitation of many large studies is the timidity of their analyses; some studies, despite having data on a considerable number of variables, have relied mainly on three and four-way tables of relationships, a fundamentally restricted view of what is a much wider and more interesting picture of relationships within an integrated model. An even more serious problem is that a limited analysis may present associations which in fact would be swamped by some of the more important variables not included in the analyses.

Another serious limitation of many analyses is the frequent - and usually unjustified - use of stepwise and hierarchical techniques. The first allows the analysis itself to determine the order of variable entry into that analysis, and the second (hierarchical) forces variables into the analysis in an order determined by the research worker. The grave problems of interpretation associated with both these approaches - as compared with a strict adherence to a longitudinal sequence of entry, with simultaneous entry of variables into a model at any single point in time - are dealt with in chapter 5. Yet a further limitation is that analytical techniques which collapse variable values into broader categories also serve to limit the potential power of such variables within the model as a whole.

Of the theoretical models of early attainment, Ramey and Gallagher (1975) offer a fairly comprehensive set of relationships from which a prediction of early attainment can be derived, given the needed data collection. They see socio-cultural influences as leading to factors such as family economics, parental education, the family's social status and parental expectations. These in turn create family patterns expressed in the form of the quality of nutrition and biomedical care, the home organisation, the language environment and social interactions. Both socio-cultural influences and family patterns help determine the child's status, including its self-concept, cognitive skills, language development, need for achievement and style of social interaction. This model is much bolder than most; its limitation is that it is totally linear, taking no account of the possibility that the child's cognitive or other characteristics may in turn influence parental expectations and other aspects of the family and home environments.

Given the variety of contributors to early attainment which have been discussed in the preceding sections of this chapter, it is possible to formulate a general theoretical model, using latent variables as broad descriptors of what are in reality a multiplicity of individual variables within each latent variable.

The model, portrayed in figure 2 overleaf, is conceptualised within four time spans, with variables in each phase being seen as having reciprocal influences on one another. Clearly the sequence of action and reaction within any interactive phase is likely to occur at frequent intervals and on a continuing basis through successive micro-periods of development, but for the sake of clarity the model is confined within the conceptual bounds of the four separate phases.

As presented here the relationships are deceptively simple. The reality is that there are not only a series of overlapping phases of development, but also spurts and delays caused both by the surrounding environment and by maturational processes within the child himself. For each latent variable there are a large number of measurable characteristics of a similar nature, as well as varying interactions with particular characteristics or phenomena within other latent variable groupings. What is important here is that the model attempts to portray the wide range of major influences which predict to early school attainment.

The problem of building such a model around actual data would be a formidable one and has yet to be tackled in a comprehensive way, using sufficiently large numbers and a sufficiently comprehensive range of variables. Inevitably there would be compromises, especially in regard to the number of variables which could be assessed on any child, family, school and community over an extended period of time. The selection of the most crucial variables would clearly be

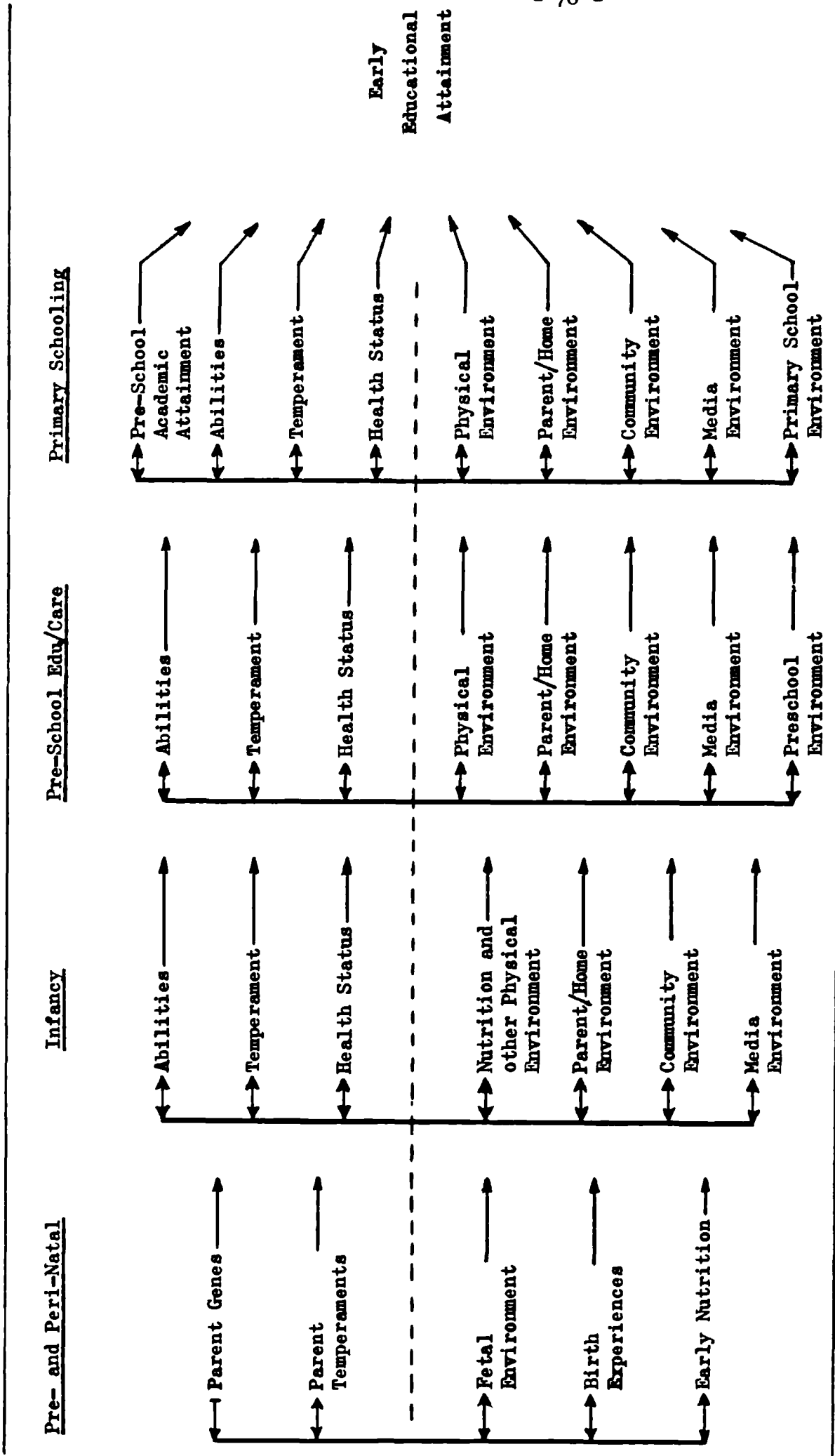


Figure 2. Theoretical model of contributors to early educational attainment

linked to the theoretical positions of those building the model. The nature of the statistical analyses of the data might be the most important part of any such model, since existing attempts to relate large numbers of variables to each other tend to err either through over-simplification, or, if using more advanced techniques, they rely on inappropriate applications and rules whose origins may lie more in the criteria for agricultural experiments and the physical sciences than in those for 'soft' psychological, educational and social models.

The basic model which is to be developed in the course of this study is limited by the fact that both the data gathering and experimental programme are the responsibility of a single person. Thus the variety of data which can be used in the model is of necessity restricted in comparison to what would be needed in terms of the broad criteria set out in this section. It is intended however to present statistical analyses which will offer a more wide-ranging interpretation of the longitudinal data than is usually met with in studies of this kind.

3.00 Intervening in Disadvantage:  
home and pre-school

The wealth of intervention initiatives which took place in the United Kingdom and United States in the late 1960's and early 1970's was followed by a period of disenchantment, perhaps in large part due to the apparent failure of many of these initiatives. Even successful initiatives tended to have only limited and short-term effects on the skills and attainments of the children who were their target.

This chapter commences with an analysis of the attempts made to define educational disadvantage in terms of social disadvantage, and points to the confusion which has arisen as a result. This confusion, though understandable, has had a considerable influence on the approach to intervention itself: should it be based on children identified as poor, or on children identified as facing early educational backwardness? While both the home and the surrounding social environments are easily identified as part of a situation of disadvantage, the extent to which the schools themselves serve to hold back the attainments of some of their children has also merited research analysis.

The next section offers an extended review of a large number of intervention programmes in the field of early childhood education. Based on the conviction that the disturbing level of educational failure in modern society has its origins in the pre-school and early school years, a massive American programme known as Head Start and a comparatively large British programme centred on Educational Priority Area schools were each launched to reduce this failure. A great many other institutional initiatives were also undertaken in these and other Western societies. A smaller number of studies were carried out which aimed at influencing the child within the home environment, or at altering the home environment itself with a view to bringing about change in the child. The interpretation of intervention differed widely between programmes, ranging from structured academic activities to the provision of totally informal settings containing a variety of educational toys and other equipment.

Among the conceptual questions raised in many of the programmes is the unresolved debate between the 'structured' and 'discovery' approaches to early childhood education (or between amalgams of the two approaches), the even larger debate - which may now be on its way to resolution - of whether to focus on the parent and home or on the institution, and the decision as to whether the

content of intervention programmes should be oriented towards increasing cognitive levels, communication skills, academic attainment or social skills.

The two subsequent sections review critically many of the underlying concepts and perspectives in the area of intervention. The first of these sections discusses the topic of action research, together with some other problematical issues in regard to the goals and strategies of intervention research; the problems which arise with design, and the even more intractable problems of sampling, are also discussed in this section; its concluding pages cite three major reviews of pre-school intervention research, with some of the new perspectives which they offer.

The second of these sections looks at three other aspects of intervention research: the evaluation of programmes and the dominant criterion of I.Q. change, with the consequences that this focus on I.Q. has had on the design and content of much intervention work; the neglected area of cost-benefit accounting in reporting on major intervention programmes; and the difficult final stage of any research, namely its dissemination and, hopefully in the case of intervention studies, its translation into public policy initiatives or into the field of public debate between policy-makers and research innovators.

The last section of the chapter summarises the principal concepts underlying the intervention programme devised for this study and concludes by setting out three formal hypotheses.

### 3.10 Educational and social disadvantage

The concept of educational intervention has undergone various changes in terminology but in essence this concept is little different from that which led to the founding of the pre-school movement for disadvantaged children nearly two centuries ago. Definitions such as compensatory education and equality of educational opportunity have enjoyed wide popularity because they express the sentiments of wanting to 'compensate' the socially disadvantaged child for what he or she is thought to lack in preparation for early schooling, or because the idea of equal access to education is clearly acceptable as a policy slogan. The increasing use of these terms has however brought with it increasing criticism of their theoretical basis: it is now claimed to be offensive and untrue to suggest that children living in poverty should have to be 'compensated' for what their parents have given them; the belief that equal access to education is an adequate goal has also been questioned, on the grounds that equal access implies an equal ability to profit from what is provided - something which is clearly improbable.

Educational intervention as practised in Western societies is related more closely to social considerations than to educational ones. While schools offer a variety of remedial strategies to assist children who are identified as slow learners - usually only after a few years of relative failure - intervention aimed at pre-empting this relative failure is usually based on the social background of the children. Banks (1968) points to some of the contentious issues, including the disputes between elitist and populist views of education and the powerlessness of poorer parents faced with a school system directed by advantaged people. Wall (1975) speaks of the contrast between the theory of educable capacity, linked to the environmental and genetic contributions to that capacity, and the theory of cultural deprivation in which the parental and other formative influences may not be sufficient to prepare children for the demands of a dominant middle class culture.

While chapter 2 has discussed most of the known contributors to educational attainment, this section will deal only with those environmental contributors which are thought to be open to change or intervention, namely the home, the wider social environment and the school.

### 3.11 The home environment

A great deal of research, some of which is referred to elsewhere in this study, has identified the parent or home environment as crucially important in



the early educational preparation of the child; more recent research, summarised by Bronfenbrenner (1974b, 1979) suggests that intervention strategies based on the parent may have more long-term success than strategies based on the school or pre-school institution.

Two studies on large numbers of children have concluded that the home environment is more powerful than that of the school. Peaker (1976b) examined data from 3,000 children in preparation for the Plowden Report (1967). Using path analysis, but relying on stepwise regression to determine the order of variable entry, he concluded that parent variables, such as parent interest and attitudes, were more powerful predictors of educational attainment in primary school than the home circumstances or the school environment. The re-analysis by Coleman (1975) of the IEA studies (Comber and Keeves, 1973, and Thorndike, 1973, *inter alia*) has already been referred to earlier. Coleman too used path analysis, and concluded that home variables predicted more of the variance in reading than did school or instructional variables, across a range of national samples.

Both studies suffer from the limitation that their data are essentially cross-sectional rather than longitudinal. An equally serious limitation is that the finding of these relationships is not necessarily evidence that the identified contributors can be modified to bring about educational improvement.

Studies reviewed by Caldwell (1976) offer a more substantial though not conclusive basis for identifying the specific contributions of the home to early development. While the part played by socio-economic variables is recognised, it is the patterns of maternal behaviour in the rearing of the child which have been found to be of major importance in determining the level of development of the pre-school child.

### 3.12 The wider social environment

Many parent and home variables have their origins in the norms, values and practices current within the surrounding social environment, and to that extent the educational importance of this environment cannot be minimised.

Major national studies in the United Kingdom, with Douglas et al (1969), Davie, Butler and Goldstein (1972) and the Newsom Report (1963) offering only a few of the large number of reports on these and other studies, all show the close relationship between social disadvantage and educational backwardness. The many indicators measured within the national cohorts point to the links between poverty, ill health, limited qualifications in the parents, lack of parental interest in schooling, and a degree of school failure in many of the children.

The more subtle contributions to performance arising from crowded living conditions, reduced educational opportunities even within a system based on the concept of equality within the State sector, and more distant but still significant contributors such as birth weight, have also been teased out of the cohort data. The three national cohort studies (Douglas et al, *ibid*, Davie, Butler and Goldstein, *ibid*, and Chamberlain et al, 1975) have the particular value that they are all longitudinal in design and although relying in large measure on cross-sectional analyses they are also now being opened to the more fundamental interpretations of outcomes over time. The description by the Newsom Report (*ibid*) of what the authors termed 'half our future' is almost as appropriate as a description of educational disadvantage in the United Kingdom today as it was at the time it was written.

Kelsall and Kelsall (1971) review the whole field of educational opportunity and attainment in relation to social disadvantage. In addition to the better known cognitive and personality factors within the child there are factors such as the effects of cumulative deficit building on early deficit, a limited attention span, disciplinary problems and the probability of unmet physical needs such as hunger. Within the community itself other factors such as ethos and attitudes towards education are important. Within the school there are a host of potentially important factors such as the level of resources in poorer and richer areas, school procedures which may often be geared more to the needs of middle class parents and children, and teacher characteristics and their expectations based on what is thought typical of working class children.

The difficulty of defining disadvantage in educational terms has already been referred to in the introduction to this section. Wolfson (1976) considers that the debate on the issue centres on two different interpretations: one is that of 'deficit', seeing the child as deprived of early stimulation and language; the other is the 'difference' view, in which cultural variations are recognised as putting the child at a disadvantage, but are seen to have their own validity even if they are not accepted by the school.

The question of whether educational and social disadvantage is a cyclical phenomenon, repeating itself in successive generations of the same families, has been a matter of considerable interest in the United Kingdom, both as a matter of concern for State policy and as an issue of fundamental research interest. Rutter and Madge (1976) examine the evidence and conclude that while there is much to support this concept in relation to a variety of child and family characteristics, there is little clarity as to why or how a considerable number of families escape from the cycle of deprivation. In every generation there are many families whose attainment levels are well above those of the parent families, and even within families children may differ considerably in later outcomes. The authors consider that it is a delusion to think that nothing short of massive

social change can influence cycles of disadvantage. What is needed is research into how to bring about discontinuities in these cycles.

A somewhat different and more problematic approach to the question of educational and social disadvantage is offered by Jencks et al (1972), whose study of inequality in adult life - based on a large volume of data and sophisticated analyses - concludes that factors such as chance or 'luck' contribute more to lifetime income than do home or school differences or inequality in educational provision. Jencks and his co-workers argue that schools have failed to equalise either short-term achievement levels or long-term levels of educational attainment and adult income; if income equality is a policy goal it should be achieved by more direct attempts to narrow the range of family incomes.

This study led to an intense academic debate; some of the commentaries are published in Harvard Educational Review (1973). The critics question inter alia the quality of the statistical analyses and the potential influence of some of the variables not included in the Jencks data. A group of Black educationists and social scientists point out (in the same volume) that the conclusions to be drawn from the Coleman, Moynihan, Jensen, Jencks and other national studies is that since the case for educational and lifetime equality has not been shown to be linked to school resources, the case can no longer be made for equality of provision. The attendant publicity and the ethnocentric nature of much social science research have been damaging because factors such as personality differences and racial discrimination have not been taken into account in the analyses. Moreover the schools which have been given the funds to offer compensatory education have not adapted their methods to the differing cultural needs of the children, being based largely on the needs of affluent white children. This group of Black professionals is particularly critical of the fact that the Jencks team applied the conclusions of path analyses based on 'native born White non-farm men' to the Black population, instead of attempting a similar analysis for the latter group.

Coleman, in the same volume, is critical of the conclusions that the unexplained part of the variance can all be attributed to unmeasured personality factors and luck. There are many other possible variables which have not been measured. Moreover it can be argued that there are important outcomes such as satisfaction and achievement which are of equal or greater importance than ultimate income levels; it is wrong to base major conclusions about educational predictors almost entirely on a dependent variable such as income.

A somewhat different study of a large sample of British men has been reported by Halsey et al (1980). Here various social and educational predictors have been examined in path analysis and other models, to determine the extent to which social class is a major contributor to both intelligence and educational

attainment. The authors find that class is still a key predictor and that the effect of various changes in the educational system have if anything profited upper class boys more than working class boys. Like Jencks and his colleagues, these authors are pessimistic about the possibility of social change occurring as a result of changes in the educational system. At each stage from primary school upwards, and with each change in the education system, merit and meritocracy are modified by class discrimination. The factor of birth remains a dominant feature of both society and the educational system.

It is interesting to note that these contrasting studies, on each side of the North Atlantic, confirm earlier evidence that American society and education are characterised by meritocracy to a much greater extent than is the case in Britain; in the latter society, social class appears to be more important than meritocratic considerations in determining progress.

Two radical studies, in Britain and the U.S.A., take as their theme the argument that disadvantaged children suffer educationally because they are different rather than being in deficit. Keddie (1973) examines what she terms the myth of cultural deprivation, while Ginsburg (1972) offers supportive evidence on the same theme. Both authors cite extensive comments by workers such as Labov on the alleged limitations in the language spoken by non-standard English speakers, and quote other studies pointing to major cultural differences in the understanding of education and the use made of it.

While both authors, and Keddie's colleagues in her edited volume, criticise the social systems within which disadvantaged children's education suffers, they do not appear to recognise the problems of ensuring that children from a minority social group gain the skills needed to compete for jobs within a dominant majority group culture. However adequate the minority language structures and the more limited or more practical educational attainments of minority children may be within their own cultural milieu, it is an historical truism that at the centre of every civilisation or metropolitan society, the minority group members who succeed are those who can adapt to the majority group culture, even if they also enrich that culture by introducing minority group customs or ideas.

Although it is not beyond reasonable possibility for the members of a socially disadvantaged minority group within a population to conform to some degree to the demands of the dominant cultural group, the problem of adaptation faced by minority groups identifiable by ethnic characteristics clearly different from those of the majority group is far more intractable. If members of a social minority group can be freely accepted for employment and advancement within the majority group once they meet the minimum linguistic and educational criteria of the dominant group, the question of discrimination becomes one of more subtle class and other distinctions, often hard to establish with any cer-

tainty. For groups faced with racial discrimination, even the attainment of the linguistic and educational criteria required by the dominant culture are often insufficient to ensure reasonable acceptance. Thus it is at the level of ethnic difference, or a total difference in life-styles such as that exhibited by the Romany travellers, that problems of gross educational or occupational disadvantage become matters of overwhelming concern.

### 3.13 The school environment

In view of the focus of the study to be reported here, on disadvantaged children in inner city Nursery and infant classes, the report by Chazan and Williams (1978), on the effects of deprivation in relation to infant schooling and performance at the end of infant school, is of interest. The authors followed 700 children in three centres - London, Swansea and Cardiff - and used a variety of tests to assess the children's language, mathematics and other skills in the third year at infant school. While children from the middle class areas performed, as expected, well above the level of children from working class areas, there was a noticeable difference between the scores of children in settled working class areas and children in what were defined as deprived working class areas.

The results of intelligence measures suggested that children in deprived areas were under-functioning relative to their non-verbal intelligence; a considerable minority of poor readers in this sample had average or above average cognitive test scores. A further division of the children from the deprived areas into those having what was considered a deprived home background, and those with a less deprived home background, showed differences in the father's employment and in the mother's contact with the schools, as also in the children's language and maths test scores. However there was no significant difference between the groups in the general information test.

A surprising difference was that the cumulative deficit phenomenon, referred to earlier in this section, was not manifested in a widening of the language differences between the two deprived sub-groups over the years of the study, suggesting that the infant schools concerned had succeeded to some extent in minimising the effects of serious disadvantage. It was also noted, in an examination of the children's emotional development, that nearly 25 per cent of the children entering infant school showed behavioural problems, while 13 per cent needed special attention. Both this and an earlier study (Chazan et al, 1971) found wide differences in the parents' preparation of their children for school, their attitudes to it once the children were enrolled, their involvement in school activities and their methods of behavioural control over the children.

While the Chazan and Williams study emphasises the interlinking between the school, home and social environment, in relation to educational outcomes, within the school itself there are various socio-educational characteristics which contribute to the performance of children. Two of these concern teacher expectations (and the consequent motivation of the child), and the ethos within the school.

The experimental intervention by Rosenthal and Jacobson (1968), in which teachers were fed with fictitious information on the high potential of a number of their children, was fully reported by these authors and equally fully criticised by a considerable number of research workers on the grounds that the findings on the 'dramatic effects' of the teacher expectations were not borne out by the tests reportedly used, by the statistical analyses nor by subsequent replications. However the usefulness of this study is that it led to a greater focus on classroom situations in which genuine teacher expectations have been shown to influence children's academic performance (for example, Grieger and Saavedra, 1971, and Seaver, 1973). This is of particular importance in situations where children's working class accents and culturally related behaviours can lead teachers to have lower expectations of the children's academic performance than would be warranted by an assessment of their intelligence. A pertinent example of this, in the United Kingdom context, was the finding that the introduction of 11-plus ability tests led to a noticeably higher proportion of working class children being admitted to Grammar School, compared with the previous situation in which teachers' expectations were the principal criterion for selection (Floud et al, 1957).

This theme is related to the much wider theme tackled by Rutter et al (1979) in which a research team followed 2,000 children in 12 secondary schools over a period of years. The authors found that the school ethos as defined by careful observational schedules and other methods of assessment was a major contributor to both educational and behavioural outcomes, with some schools achieving well above the level of other schools, although all the schools were located in disadvantaged inner urban areas in London. Their detailed study has been criticised strongly by Tizard et al (1980b) and Heath and Clifford (1980), but more perceptively reviewed by Wragg et al (1980). The burden of the criticisms was that insufficient attention was paid in the analyses and conclusions to the importance of the social environments within which the different schools were sited. The critics also argued that although Rutter and his colleagues had concentrated on the 15,000 hours spent in school, the influence of the 70,000 hours spent outside the school over the same period had not been taken into account. The interpretation of ethos was also questioned.

Perhaps the most crucial consideration in assessing the importance of the school in relation to educational intervention is the extent to which school

itself offers a moderately equal opportunity to most of the children who attend. The study by Halsey et al (1980) has already been cited. They found that in the United Kingdom the social class determinants of attainment are so strong as to outweigh the educational ones; however they also note gross disparities in the educational provision itself.

Another study of major importance which examined the question of school equality was that of Coleman et al (1966). The most surprising finding of Coleman et al was that there was virtually no correlation between pupil performance and school facilities. Data was gathered from 600,000 American school children and the schools they attended.

There have been a number of criticisms of the findings, inter alia that the methods for statistically separating out the influence of home background from that of school factors underestimated the school characteristics; that school expenditure per pupil was based on districts rather than individual schools; and that the main yardstick of performance was a verbal one, rather than academic attainment, reflecting thereby home and family influences rather than school factors.

A group of papers edited by Mosteller and Moynihan (1972) reviewed the Coleman Report in some depth. The authors pointed to the difference between aiming at concern over equality of outcome, as in Coleman's study, at equality of access and opportunity for optimal development. It was of interest to note that the child's sense of control over his or her environment correlated more strongly with educational achievement than did any of the family or school variables. Otis Duncan, writing in the same volume, showed that although it would be preferable to use experimental data, the Coleman survey data indicated that a \$3,800 difference in income between Blacks and Whites was largely accounted for by:

Years of schooling	\$ 520
Family background differentials	\$ 940
Occupational difference	\$ 830

Duncan concluded that fair employment practices could achieve more for the Blacks than could longer schooling. This is a questionable argument, however, since the years of accredited schooling, particularly in American society where credentials are a principal determinant of job opportunities, may be essential for most occupational access. Years of schooling, in turn, may have as much to do with family background as it does with the school environment itself.

While the Coleman et al study has often been used to suggest that schools do not count for much, compared with home and social variables, the limitations of their analyses suggest that while the importance of the latter cannot be denied, the balance between the influence of the home and social background and that of the schools may not be as heavily loaded against the schools as has been claimed.

### 3.20 The reality of intervention

A large part of chapter 2 describes the conceptual framework within which the education of the young child has been visualised by educationists and others concerned with development in early childhood. The foregoing section (3.10) sets out the background of social and economic disadvantage within pre-school intervention has come to be regarded as necessary for many children, on the grounds that the level of stimulation in their homes is considered to be inadequate for preparing the child for the linguistic, social and cognitive demands of school.

The present section will review selectively a variety of intervention initiatives in the United States, United Kingdom and in some other European countries, and will report some of the judgements of researchers and other commentators on the effectiveness of the different programmes. Critical comments on the goals, evaluation and other aspects of these programmes will be reserved mainly for subsequent sections.

Perhaps one of the most profound statements yet made on child rearing and education in the developed countries of both the Eastern and Western blocs was that expressed by Bronfenbrenner (1974a) in his review of what he termed the two worlds of childhood. He presents the contrast between a Western world in which the interests of the individual are paramount - so much so that there is an increasing degree of age segregation and alienation of the young, with a reduced concern about child-rearing and development - and an East European world in which the development of the social group and its social cohesion and awareness are paramount, with a loss of individual liberty and expression but a compensating intense focus on children and their growth and welfare.

Going beyond Bronfenbrenner's thesis, it may be argued that since individualism is a dominant characteristic of Western society, intervention may have become a necessary strategy to minimise the damaging effects of a social system in which laissez faire economics play a stronger part in determining progress than do the economics of social engineering. In other words, the very Western freedom which has until now ensured rapid economic growth has also brought with it a considerable degree of individual and group failure, as measured by poverty, unemployment, communal disintegration and the resulting social distress; it is possible to argue that it is these factors which are at the heart of much parental inability to cope with the demands of early child-rearing and early stimulation.

Whether or not this is a tenable hypothesis, educational failure is a major problem in Western society. Both policy makers and research groups have focused



a great deal of attention on the potential of pre-school intervention programmes for improving the level of functioning of children living in disadvantaged circumstances. How this functioning is defined and the attempts made to alter it are the principal themes of this section.

### 3.21 Head Start: the American dream

Head Start was a typically American response to what was seen as a serious socio-educational and socio-economic problem. For American society the link between education and the economy has always been viewed as a close one. Studies such as those of Butts and Cremin (1953), Berg (1970) and Gumbert and Spring (1974) describe the close interweaving of educational opportunity, credentialism and employment. In the late 1950s the Soviet launching of Sputnik, well ahead of the equivalent American manned satellite launches, brought a radical rethinking of the goals and curricula of American education. While part of the attention was focused on education at high school and college, there was also serious concern over the level of school failure. A 1963 survey found that an average of 28 per cent of White youngsters could be classed as school dropouts, with the figure for Black youngsters rising to 57 per cent.

Inevitably the attention turned to earlier predictors of school failure, to the great numbers of children who were reaching school with little language and few of the normal competences on which the first few years of schooling need to rely. It was in this highly idealistic era of Presidents Kennedy and Johnson that imaginative but hastily constructed proposals were made for a massive intervention effort focused on the pre-school child. The Federally funded programme was given the name of Head Start. There was no single model for the programme, since it was essentially a very large initiative in which school districts and other educational authorities could apply for grants to carry out a variety of local programmes. While these were all educationally based, many of them included health and nutritional intervention.

A great many documents describe the early years and early findings of the Head Start programme. Kelsall and Kelsall (1971), Zigler (1972), Shipman (1973) and Austin (1976) offer reviews of the research on what has been described as the legacy of the war on poverty. Deasey (1978) describes the programme objectives as set out by a Federal panel of experts. The children's physical health and physical abilities were seen as the first objective - 21 per cent of Head Start children were subsequently identified as having major medical problems. Social and emotional development was seen as the second goal, with cognitive development put into third place. The remaining four goals were aimed at fostering motivation and social relationships, improved attitudes towards society

and an increased self-esteem in both the children and their families. Deasey points out that subsequent research found Head Start to have had higher effects on motivation than it did on ability test performance.

Austin (ibid) shows that before the advent of Head Start in 1965, pre-school education in the United States was for the most part provided for middle class children and was thus oriented towards fostering the development of social and emotional skills. In contrast Head Start set out to provide a comprehensive programme for the disadvantaged children, including social services, health, nutrition and education; the first three were included because it was realised that poor, sick and hungry children were unlikely to learn well.

The most difficult part of this programme, according to Austin, was the educational component. The only basis for this were the theories of Decroly, Montessori, Piaget, Gesell, Hunt and Bloom; few of these had any clear practical programmes to offer to stimulate the children, intellectually or academically. Thus new programmes had to be devised and tested and it was not surprising that there should not have been any massive breakthroughs in the first ten years. What was shown was that there was a specificity of effect, linked directly to the nature of the programme.

A more critical view is offered by Clarke and Clarke (1976) who in a chapter entitled 'Some contrived experiments' argue that Head Start was founded on at least four erroneous principles: the use mainly of very short programmes; the common use of a play-school approach, offering the child little education and little cognitive activity; treating the child out of her or his home context; and failing to reinforce any gains by repeating the enrichment when the child moved into school.

A massive and costly research evaluation by Westinghouse Corporation in association with Ohio University (described in Cicirelli et al, 1969) yielded the first major report on the impact of Head Start. The report was critical of the programme and the general burden of the findings was that Head Start had not been effective. This study, more than any other, contributed to a measure of deep disillusionment, political and administrative, with the programme. Although Head Start was never abandoned totally and is in fact still in existence today, the level of funding was sharply reduced as a result of what is known as the Westinhouse/Ohio report.

The only finding which has never been questioned is that children's health benefited considerably as a result of the medical treatment and daily lunches given to programme participants.

Perhaps the biggest flaw in Head Start was that funds were poured in wherever an educational district or State identified a need, whether or not there were clear goals in the proposed programmes. There were few demands for accoun-

tability, and there seems little doubt that a great part of the early funding was spent more with enthusiasm than with any evaluation of effectiveness in mind. Much use was made of para-professionals in the form of mothers and teenagers from the local communities, but their training was patchy or non-existent. In a situation such as this it was easy for a high-powered team of researchers to find the case for Head Start unproven, particularly as the criteria were those of improved cognitive ability and affective development.

At the same time it should be noted that the statistical methods used by Westinghouse/Ohio in finding against Head Start have been severely castigated by Campbell and Erlebacher (1970), who report that the evaluators' incorrect application of matching and analysis of covariance techniques in an ex post facto situation could in fact have proven that Head Start was positively harmful. It was not the insistence of hard-headed politicians on a proper evaluation which created the damaging findings, according to Campbell and Erlebacher, but rather the inadequacies of the social science methodological community which was not ready for this massive task.

Two other reviewers (Bissell and Smith, 1970) have pointed out that even on the basis of the Westinghouse/Ohio survey, Head Start reduced the cognitive gap by between a quarter and a half standard deviation for the total sample and by close to half a standard deviation for the Black sample. (The Westinghouse/Ohio team applied the single decision rule of accepting only a difference greater than half a standard deviation as 'important'.)

While individual research teams, based on various universities or other research centres, continued to report that their own Head Start programmes had been successful - apparently because they had clearly defined goals and, frequently, clearly structured methods - it was not until a decade later that a consortium of 12 of the most prominent university participants in the early Head Start work set up a study of the progress and achievements of the intervention and control children who had originally been identified in the 12 pre-school samples, some of them very large samples. The collected findings, reported in Lazar et al (1977) and Lazar and Darlington (1978), show that on every major criterion the relatively large samples of intervention children were ahead of their control peers. There was less class retention - the practice followed in the United States with poor achievers - and less school dropout, and there were improved reading and mathematics scores.

Other intervention programmes linked to Head Start were Planned Variation (reported by Bissell, 1973) and Follow Through (Bissell, *ibid*, Abt Associates, 1977, House et al, 1978 and Anderson et al, 1978). As with the original Head Start evaluation, the Follow Through evaluations were the subject of bitter academic debate, the flavour of which can be sampled from the two latter references. The Follow Through programme, as its name implies, was a programme focused on

Head Start children immediately after their entry into the elementary grades of school. Both Planned Variation in Head Start and the Follow Through programmes offered more educationally oriented and more structured activities for the children that Head Start did, and both had a measure of success.

The overall lesson from the various Federally funded initiatives is that in so far as they provided clearly definable programmes, with specific goals, formal evaluation pre and post hoc and reasonably identified control groups - while recognising the difficulty of obtaining totally comparable sets of control children - moderate to good effects have been obtained, particularly in the area of academic attainment. On the other hand cognitive growth resulting from these programmes has generally been short-lived, returning to levels not much above the expected levels within one to three years after the ending of the programme. It is more difficult to assess results in terms of affective development, since this is less clearly definable and is open to a variety of interpretations ranging from expressed self-confidence to observers' ratings of the children's motivation within specific task situations.

### 3.22 Educational Priority Areas: the British compromise

Despite the urgent recommendations of the Plowden Report (1967), proposing Government action to tackle the increasing problem of early educational failure in Britain, the first reports on the problems being experienced with the Head Start programmes in the U.S.A. led to a far more cautious approach to Government-funded educational intervention in the United Kingdom. The awareness of large scale educational failure was not as strong in this country as it was in the United States, and the news of Sputnik did not arouse the same intense educational self-questioning that it did in the country which had considered itself to be in the forefront of world technical progress.

Faced with moderate political pressure for some educational action, but also aware of the research caveat that the spending of vast sums of money could be wasted if not properly directed, it was decided to compromise with a set of limited budget initiatives in various parts of the United Kingdom. The initiatives were all focused on what had been defined by Plowden as Educational Priority Areas, areas identified by various social indicators as having an unusually high proportion of poor and very poor children living in them. While these initiatives were intended as a small scale beginning to what would ultimately be a large programme of State intervention, in fact the E.P.A. programme as such was never developed further and was followed simply by the continuation of existing arrangements whereby teachers working at schools in E.P.A. areas were paid an additional allowance, while class sizes were reduced to enable more attention

to be given to the E.P.A. children. In effect the Plowden strategy for higher salaries and smaller classes in E.P.A. schools remained the only significant innovation following the ending of the E.P.A. research programme.

Some of the more important features of this research programme are described in Halsey (1972), Lines and Widlake (1971), Smith, G. (1975), and Armstrong and Brown (1979), *inter alia*. The initiatives included various language oriented intervention programmes, a mathematically oriented programme and a home intervention programme. In contrast to the large scale of some of the Head Start experimental programmes, the E.P.A. ventures involved relatively small samples of children, carefully assessed. It would be over-simplifying the differences between the American and British approaches to describe them as a contrast between a 'hard sell' emphasis on desirable activities, and a 'softly softly' approach of providing an enriched classroom environment, but there are hints of this contrast in authors' descriptions of the major programmes in the two countries.

In a preface to the report on the Red House (home intervention) Project by Smith, G. (*ibid*) - in the final study of the four-volume account of the E.P.A. programme - Halsey concurs with the conclusions of Smith and his colleagues that the main body of recommendations on positive discrimination, put forward in Volume 1 by Halsey (1972), have not been taken up and have in fact been left 'badly becalmed'. Smith himself concludes that the evidence either for success or failure has scarcely proved robust enough to justify the sudden lurches of national educational policy (away from the original political intentions in favour of positive discrimination).

The results of a subsequent study by Armstrong and Brown (1979) were seen by these authors as particularly disappointing. They examined the intervention and control children identified in the Red House Project and found that there was no discernible difference between the groups on either cognitive or reading measures. However they did find an improved attitude towards education on the part of the intervention parents and some improvement in parental behaviours, such as time spent with the child and the number of out of school activities, compared to control parents.

The E.P.A. concept itself has been the subject of critical scrutiny. Acland (1971), commenting on the actions taken by the Department of Education and Science to implement the recommendations of the Plowden Report (1967), shows that the recommendations have meant an increase in resources and in teacher quotas and salaries within geographically defined areas of deprivation. He questions this approach, which is based on the assumption that there is a narrow range of ability within each school, and that priority area funding on a geographical basis will therefore benefit most children in need. The reality is that the range of performance in most schools overlaps very considerably with the ranges in other

schools. Resources should be allocated on the basis of the proportion of under-achieving children within each school.

Acland also doubts whether the E.P.A.'s loose social definition of the deprived child and the diffuse Plowden aims, such as improving children's self-concept, developing their understanding of the community they live in, and using school as a forum where the community can realise its identity, are of much immediate relevance to the educationally backward child. It seems necessary to take a narrower view of compensation, basing it on the need to improve a child's academic achievement - since this is fundamental for later occupational careers.

The inadequacies of Plowden's area or school-based approach to identifying children in need of extra help were highlighted in research undertaken by Ferguson et al (1971) into the use of school characteristics. The authors suggest that infant teachers could preferably identify children whose poor home conditions and inadequate range of prerequisite skills for early education point to the need for a direct approach to the problem. A recent review by Mortimore and Blackstone (1982) cites a number of studies which have also questioned the Plowden criteria for helping children at risk, but these two authors consider that the alternative of allocating resources to individual children may have the effect of labelling them. They also cite other administrative snags in funding special programmes for designated disadvantaged children.

### 3.23 Three landmark studies

While it is inevitable that the results of comprehensive and multi-centre national intervention programmes will be subject to a great deal of politically motivated evaluation, frequently overshadowing the independent 'hard' research perspectives, successful individual studies can be more useful in focusing attention on particular approaches to intervention. Three of these deserve mention for their unique features.

In what is known as the Milwaukee Project, reported in Garber and Heber (1973) and Heber and Garber (1975), a sample of 40 Black infants were drawn at a few months of age from a seriously deprived urban environment and randomly allocated into experimental and control groups. The experimental children were given an intensive full-time nursery experience, with highly structured activities, for a period of several years. Much of the work with the babies and growing children was carried out on a one-to-one basis. The children's mothers were given job training and provided with employment during the same extended period. Efforts were also made to guide the mothers' home-making and child-rearing skills. By the time the children were  $5\frac{1}{2}$  years of age their mean I.Q. score was 30 points above that of the control children.

Other improvements also took place; the authors note the particular finding that children in the experimental group were more likely to provide information for their mothers in interactional situations, and more likely to stimulate and question their mothers, than were the control children in a similar dyadic situation. Criticisms of this project are described in a footnote on page 133.

The question of the cost-effectiveness of the Milwaukee programme will be discussed in a later section of this chapter.

A very different approach was that used for a series of pre-school programmes which made up the Ypsilanti Perry Preschool Project. These programmes are described in a number of books and other publications, including Weikart et al (1970), Lambie et al (1974), Weikart, Epstein et al (1978) and Weikart, Bond et al (1978). The project covered over 120 children, approximately half intervention and the remainder controls. The children entered the study in five successive overlapping waves, starting in the early 1960s, and were followed through most of their school years. All the children, experimental and control, were drawn from families with a low socio-economic status and were selected on the basis of having an initial intelligence score between 50 and 88; this was followed by sequential ordering of the children according to IQ scores and then dividing them into odd/even groups, with further S.E.S. and sex equalisation to ensure absolute parity between treatment and control groups.

The intervention children were given an intensive two-year pre-school experience, with a child-teacher ratio of 6 to 1. The programme included weekly home visits. Analysis of covariance showed that while the measures of I.Q. indicated an initial spurt for the experimental children, this advantage almost disappeared over time; however a slightly higher level of academic achievement at the end of grade 1 (compared to control children) gradually increased over the years. The difference between the groups was wider than one grade at the end of grade 8, although both groups were still well below the national norm. Academic achievement was measured in the form of reading, language and arithmetic scores.

Although in many ways this was an outstanding study and produced results of considerable importance - the cost-benefit analysis described in section 3.40 is of particular value in reflecting on the value of pre-school intervention - a disappointing omission from the study is any real consideration of the possible importance of the parent and home visits.

In view of the focus of later chapters of the present study on the design, execution and analysis of a parent intervention project, it is worth considering in some detail the omission of the parent factor from the analyses of the Ypsilanti Project.

There is no indication that the research team regarded the home intervention part of the programme as particularly important, although teachers said that the

home visits had helped them to get to know the children and parents better. For whatever reason, the statistical contribution of the home visits to the cognitive and academic outcomes was not experimentally evaluated, although the five treatment waves could have enabled one or two waves to be given only the pre-school treatment and no home visits, or indeed vice versa. Equally, the one instrument which could have been used to assess the home behavioural environment was only constructed during the programme and administered once only to every parent, four years after the start of the initial programme.

The potential value of this single assessment of home behaviours, which might still have indicated important predictive differences across waves and treatment groups - the fact that it was measured in 1966 gave it a relatively different time relationship with each wave - was unfortunately reduced even further by a statistical procedure in which those portions of the variance of home behaviours associated with treatment group and wave were removed, leaving only the residual scores on this instrument for use in the analyses.

The differences in the home behavioural variables were complicated by the fact that 31 per cent of the control group children had working mothers, compared to only 9 per cent of the intervention group, and it is fully understandable that removal of that part of the home behavioural variance attributable to the group factor - which was likely to favour the control group if the working mothers were more upward mobile than those staying at home - should have been considered. However this strategy also served to remove any possible differences in the interaction of groups and waves - differences which could have been due to intervention mothers in the later waves having by then improved their child-rearing behaviours as a result of the weekly home visits. (The home behavioural items showing any relationship to child age had already been removed prior to deriving two rotated principal components on the remaining behavioural items.)

The authors' assessments of the relative importance of the major variable groups in predicting academic achievement in each of the first four school years are of interest. The academic achievement levels in the various post-treatment samples were regressed on treatment group, child entering characteristics (a measure of initial intelligence) and a combined home environment variable which included the home behaviours principle components, a demographic variable and a composite of maternal attitudes. Treatment group effects were significant though relatively limited (but increased to a unique 5 per cent when mother's employment was entered into the model), child characteristic effects were relatively small, with unique variance predictions ranging from 4 to 6 per cent for grades two to four and 15 per cent for first grade, while the home environment combination remained consistently high, ranging between 15 and 19 per cent unique variance prediction over the four years. Although a good part of this latter variance may be due to the demographic variables (particularly social class



differences within the socially deprived population), the results do suggest that the importance of home environment could have merited fuller emphasis in the study.

With such a high contribution to academic achievement - far above those of child intelligence or pre-school treatment - there appears to be a considerable potential for intervention to alter the home behaviour segment of the overall home environment variable, in order to determine whether home behaviours form a meaningful part of the overall contribution of home environment. This could have been linked, in the Ypsilanti Project, to the further question of whether or not the home visits were in fact a significant contributor to outcome variance, alongside the contribution of the two-year pre-school institutional experience. Without evidence on possible differences in the levels of home visits and of the parallel pre-school intervention, it could even be argued that the key intervention variable may have been the home visits programme rather than that of the pre-school experience.

A small paragraph in Weikart, Epstein et al (1978) hints at the importance attached to the parent variable by a leading American specialist in early childhood education, J. McVicker Hunt. During a visit with other senior researchers to review the impressive results of this project, he is quoted as saying: "I find myself asking why these programmes have produced larger IQ gains than any other programmes of which I know. I find myself wondering if the home visits may not be the key to this gain, and I wish I knew more about them.... if they are, they promise not only to make the gains of the children more permanent but also, if the findings of Merle Karnes and the Peabody group are correct, to offer a basic enrichment to the lives of the parents themselves."

Reference has already been made in chapter 2 (section 2.55) to the third landmark study, that of Hewison and her colleagues (Tizard, Schofield and Hewison, 1982, and Hewison 1981). Although this study was not aimed at the pre-school child or its mother, its particular strength lay in the fact that it involved the primary schools concerned in large scale practical parent intervention programmes in which the parents were given the responsibility for helping their children's reading at home. Its unique features included the fact that parents were entrusted by the schools with what might be described as a para-teaching role in helping their children to improve their reading levels and in coping with the more mundane problems of the young reader - the parents being guided by daily notes from the teacher, with the parent in turn replying to inform the teacher of the progress and difficulties. The more serious reading problems were of course handled by the teachers in the classrooms.

The great strength of this initiative lies in the fact that, in contrast with most intensive intervention programmes - including the two other landmark

studies cited here - it is an extremely low cost venture and makes considerable use of the resources, initiative and good will of both the parents themselves and the existing educational service. A number of educational authorities in Britain have now adopted the methods formulated in the original Hewison study, or have developed modified forms of these strategies.

### 3.24 Institutional intervention

Most of the intervention programmes which are concerned with the pre-school child have offered institutional stimulation of the child rather than attempting to alter the home environment. The first major initiatives in the institutional field consisted of a wide and varied range of nursery schooling and kindergarten, set up especially for the children of the poor in most European and other developed countries over the past two centuries. The history of these nursery initiatives is described briefly in the previous chapter (section 2.61).

What is interesting about them is not only their focus on the institution rather than the home, but also that they introduced, possibly for the first time in recorded history, the concept of differentially encouraging supposedly less competent mothers to hand over their children for the formal development and encouragement of the children's abilities. Although a variety of play, creative and other unstructured and non-academic activities were introduced later in reaction against the over-formal nursery schooling provided by many of the early pioneers, the view still predominated that the children were being given something which they could not expect to receive from the hands of their poor parents. Governesses and other forms of sophisticated child-minding were of course often used by middle and upper class mothers, in what can now be seen as a precursor of the later liberation movement which was to enable all mothers to be freed of their permanent confinement to home and hearth throughout the child-rearing years; however it was not thought necessary that the children of advantaged parents should be enticed into nurseries, at least not in the early period of the development of this movement.

The nursery institutions were clearly aimed at compensating children for what was seen as their inadequate upbringing. The task of providing this experience was given to professionals such as teachers, although the nursery movement rapidly developed an ethos, educational philosophy and training of its own. Some of the views expressed by those who helped create this movement have been set out in the earlier historical section of this study. What is puzzling is that it was not until the second half of the 19th century that any thought appears to have been given to improving the mothers' own professional competence as a parent, and even this belated move was based more on the need to spread

sound ideas about the hygienic environment of the child rather than on any concern with other forms of child functioning. A brief reference to this particular historical development will be made in section 3.25.

Some of the special intervention programmes set up more recently for the pre-school child in Britain and the rest of Europe are described in Tizard (1974), Chazan (1978) and D.E.S. (1981). While Tizard's study is a perceptive review of the major issues in the United Kingdom's application of research ideas to the field of pre-school education, the book edited by Chazan brings together a variety of European and other perspectives in the same field. These, and Continental studies such as Osterreith et al (1977) and De Vries (1974), highlight an interesting difference between much Continental pre-school research and equivalent English studies. For many of the Continental programmes the research component is something added to a highly practical endeavour; the English programmes in contrast tend to build the practical aspects around a fairly solid research core. Thus what the English studies gain in the quality of the research they tend to lose in generalisability and immediate applicability; the reverse criticism applies to many of the Continental studies, although clearly there are exceptions in both these areas.

The editing by Osterreith et al (ibid) of four Belgian studies reflects the authors' disenchantment with large-scale centralised projects and their awareness of the political and administrative structures which need to be taken into consideration in reforming the pre-school setting; all four research teams focus on both educational and socio-cultural variables and conclude that short-term programmes cannot prevent social disadvantage interfering with school learning; pre-school programmes need to be extended into primary schooling.

In his overview of a large number of European studies, Chazan (ibid) concludes that in general researchers have not been able to show that attendance at a pre-school establishment or even exposure to a special programme has substantial long-term effects, despite limited evidence that in certain cases programme effects have been noted. The relative lack of success of compensatory programmes based principally on the school has led to a greater focus on home-based intervention and parent involvement, according to Chazan; hitherto the latter has been a neglected area for research and experiment.

A number of specific intervention studies reflect the considerable variety of endeavours which have been made in an effort to improve the pre-academic skills of young children, especially those living in situations of disadvantage.

An Irish study reported by Holland (1979) and Kellaghan (1977) provided a pre-school programme for children living in a seriously deprived urban area; the evaluation included cognitive, meta-cognitive, academic and personality assessments, as well as an assessment of the home environment. Comparison of

scores at age 8 with a control group - measured at the same age but some calendar years earlier - showed that despite the usual post-test decline, even at age 8 the I.Q. level of the intervention children was well above that of the comparison group. However there was only a marginal difference on the attainment measures. It can be noted that the project team faced a serious problem due to the high mobility of the population - and the selected sample - and it was for this reason that it was decided to use for control a parallel population of 8-year-olds who could be assessed at the time when the intervention began with the pre-school children.

There has been a very wide range of intervention programmes undertaken in the United States. Some of these have already been described. The most highly structured are those linked to the names of Bereiter and Engelmann (Bereiter, 1972, and elsewhere); these programmes involve intensive oral drill in patterns of speech and logic, with rote learning followed by a series of analagous examples of increasing difficulty. Karnes (1973) offers a useful comparison of the results of five American nursery intervention programmes. On most indicators the two highly structured programmes (Bereiter-Engelmann, followed closely by a Karnes cognitive information programme based on the skills defined in the clinical model of the Illinois Test of Psycholinguistic Abilities) showed results considerably above those of the Montessori, Traditional and Community Integration programmes. When the usual drop in cognitive gains occurred the superiority of the Bereiter-Engelmann results was replaced by those of the Karnes programme, both these groups still being slightly ahead of the other groups on measures of intelligence.

Supportive evidence for Karnes' conclusions comes in a study by Bissell (1973). She defines four categories of programmes. The first is permissive enrichment, which she describes as essentially an adaptation of adjustment-centred pre-schools designed for middle-class children, with teaching strategies being child-centred and permissive. She cites Weikart's description of these teachers as 'watching and waiting for the child's needs to determine the timing of different activities'; traditional nursery schools and a number of the early intervention programmes fit into this category.

The second category is defined as structured cognitive, where programmes foster aptitudes and attitudes directly related to the learning process rather than fostering psychological development of the whole child; the teacher takes more part in directing activities rather than merely responding to informal learning opportunities provided by the children. Many of the more clearly defined and well researched programmes fall under this heading.

The third category refers to structured informational programmes with a high degree of structure and the goal of teaching specific information; here the teachers direct children's activities in clearly delineated ways. The Bereiter-

Engelmann programme is an example of this category.

The final category is termed structured environment, in which children direct their own activities through self-instructing classroom materials within 'prepared environments'. This definition fits Montessori-type programmes.

Bissell's analyses show that the structured cognitive and structured informational type programme have significantly better results than the permissive enrichment or structured environment ones. The author's examination of a great variety of programmes also indicates that many Head Start centres were based on permissive enrichment programmes, being essentially adaptations of middle-class traditional pre-schools, with a 'watching and waiting' strategy.

While most institutional programmes have been concerned with children from two to three years upwards, Ramey et al (1977a) and Ramey et al (1977b) describe one of the relatively few initiatives concerned with children from soon after birth. In some cases children were taken into the programme at the age of six weeks, with a child/teacher ratio of 4 to 1; for older programme children this ratio was reduced to 6 to 1, from the age of three onwards. The programmes differed from most others in that they contained a large component of medical care and nutritional supplementation. Results indicated a considerable increase in I.Q. levels compared to control children. This programme in particular reflects the philosophy referred to earlier, in seeing the child of disadvantaged parents as being in need of therapeutic removal and intervention away from the home and parent.

One uniquely American approach to intervention was expressed by a six million dollar experiment carried out at the instigation of the U.S. Office of Economic Opportunity. Six companies out of 31 who submitted tenders for the application of varying intervention strategies - each company consisting of a consortium of educational experts, psychologists and sociologists - were given contracts under which they were expected to raise the attainment levels of seriously disadvantaged children by a full grade within a period of seven months. The experimental design was well controlled, and covered 18 school systems; the conditions of the contract were that payment would be based on the children's performance. Eysenck (1973) presents a critical review of the concepts and methods underlying this experiment. In several cases there were no discernible differences between the mean scores of intervention and control groups on reading and arithmetic; the maximum difference achieved between any pair of intervention and control groups was two-tenths of a grade. Eysenck regards the experiment as a massive failure, and argues that part of the reason may be that the teaching methods were not adapted to the particular needs of individual children.

An unusually comprehensive and sophisticated intervention programme was that undertaken by McKay et al (1978) in the Latin American city of Cali, in Colombia.

Over 300 children formed the intervention sample, these being drawn in successive waves at an differing ages. Their progress in a combined health, nutritional and pre-school educational programme was assessed in comparison with a control group drawn from the same socio-economic level and with another group drawn from a higher S.E.S. level. The team used Rasch methods, inter alia, to equate the evaluation instruments employed at different ages. They found a linear relationship between the growth in ability scores, relative to the control group, and the number of 'treatment periods' enjoyed by the children (ranging from 1 to 4 annual periods). Overall the authors conclude that combined intervention can limit the losses of cognitive ability in deprived children, and that these effects are persistent and measurable at the age of 8. The 'treatments' included nutritional supplementation and a full-day nursery stimulation programme.

Two recent reviews reflect the range of British intervention studies. Chazan (1975b) points out that many of these studies have suffered from the smallness of their samples; at times only one nursery school has been involved. One large intervention study was, however, carried out in Dundee, involving over 500 children from pre-school establishments and giving them a programme concentrated on specific aspects of cognitive development. The pre and post-test design included two control groups. A particular difficulty here was the loss of more than one-third of the sample. The analysis showed that many of the comparisons between experimental and control groups were in the expected direction, but few were statistically significant. A later review, D.E.S. (1981) indicates that there have been very few specific pre-school intervention studies of any size, in institutional settings, in the years since Chazan's review.

One important English pre-school initiative has been that directed by Jerome Bruner. The history and development of this initiative offers some useful pointers to the internal and external limitations on intervention research.

The Educational Research Board of the Social Sciences Research Council, following discussions with Her Majesty's Inspectorate, other officials of the Department of Education and Science, and a number of Britain's leading academics in the field of early childhood education, invited Bruner, who is one of America's foremost educational psychologists and was at the time Watts Professor of Psychology at Oxford University, to submit a research proposal encompassing programmes in four areas. The Working Group specified that these programmes should include monitored intervention at the nursery level, skill development through specific curricula, multi-disciplinary studies of the structure of curricula and invisible pedagogies in the peer-group, home and school, and socio-political studies of the community, its organisation and empowerment in relation to pre-school provision.

Bruner (1980), in one of a number of books arising out of the research in

what came to be known as the Oxford Preschool Research Project, cited this invitation, but went on to point to the difficulty of getting research findings accepted, even by those who had participated in successful intervention programmes. Following very considerable consultations with many specialists on early childhood education, including academics and a number of practitioners in the fields of day care, such as the playgroup movement, nursery education, day nurseries and child-minding, a variety of studies were set up to examine provision, practice, policies and philosophies within each of these pre-school fields. (Some of the findings have been referred to in section 2.63.)

While the studies offered profound insights on many academic and policy issues, including in particular the need for more State direction and coordination as well as more consultation with parents, practitioners and all others concerned, the work undertaken for the Project did not include the conduct or monitoring of formal intervention programmes, other than a useful study of how videotape recordings could help nursery teachers to recognise the limitations of their methods and adapt their teaching so as to have greater success in the achievement of their own goals.

Bruner's perspective thus differed considerably from that of the S.S.R.C.'s Working Group which formulated the remit. Bruner saw the fundamental issues in terms of what he calls the two large questions of whether there is sufficient pre-school provision, in quantity and quality, to secure a competent next generation; and whether there is the means in Britain for developing policies about such a question that are at once rational, compassionate and economically viable. He and his colleagues felt that the studies did offer valuable material to help answer these questions with some though not unequivocal certainty.

It can be seen that the great majority of institutional intervention programmes have concentrated on general cognitive development or, more specifically, on language and pre-reading skills. It is seldom that there is any interest in mathematics, although increased 'spatial awareness' is often advanced as one of the general cognitive goals. The Birmingham E.P.A. study (Lines and Widlake, 1971) was one of the few programmes to consider mathematical development, and as already described in section 2.20 even this programme conceived of mathematical development and evaluation in rather narrowly academic terms, based only on Piagetian criteria. The neglect of mathematics is also noticeable in most American programmes; the comprehensive review by Austin (1976) of 14 programmes shows that only two of them had any mathematical content, and even that was overshadowed by the major emphasis on reading-related goals in those programmes.

A fair number of language-oriented intervention programmes have been reported in the literature. While British attempts to introduce programmes based on

the Peabody Language Development Kit have been unpopular with teachers and relatively unsuccessful, other programmes have proved more useful. Chazan and Cox (1976) offer a comprehensive review of the theoretical issues underlying language programmes and deal with the sensitive issue of whether the language usage of sub-cultural groups is deficient or merely different. Although a good case can be made out for accepting considerable differences in usage, the Bullock Report (1975) points to the indisputable gap between the language experiences that some families provide and the linguistic demands of school education. Chazan and Cox set out what should be the goals of a compensatory language programme, namely to develop the ability and motivation of the children to use language as a means of classifying and ordering their experience and as a medium of logical thought.

American programmes with a particular language emphasis include those with psycholinguistic training - associated with the Illinois Test of Psycholinguistic Abilities - the Peabody Language Development Kit programmes, instructional dialogue and the tutorial language programme.

Blank and Solomon (1968) report on the principles underlying the last named programme. Rather than rely on a total enrichment situation, which may not be very useful for the disadvantaged pre-school child, the authors have developed a set of teaching methods which stress the usefulness of the child's previous experience in the questions put and answers evoked, with the teacher using language (but no gestures) and a variety of strategies to develop selective attention to what is said or asked; ultimately the child is encouraged to instruct the teacher and thereby becomes aware of possessing language. A consciously structured language environment is regarded by these authors as being more productive than are general language programmes, requiring only 20 minutes a day of concentrated language work with a child.

In sharp contrast with the views of Marion Blank, Joan Tough (1977) rejects the structured cognitive or language programme approach, although she recognises the difficulty faced by disadvantaged children in trying to organise their experience in such a way as to be able to communicate with others. Tough considers that children need interaction with skilled adults in order to bring out what they can do by way of organising experience and using complex language structures. The structured programmes appear to neglect the need of the child to reflect on and use its own inner knowledge to develop a wide range of language strategies.

Again one is faced with a wide contrast in approaches to intervention with the disadvantaged: on the one hand there is Blank's direct focus on the elements of language usage by disadvantage groups, and a specific set of equally direct strategies to overcome the limitations in that usage; on the other hand there is Tough's sensitive awareness of some of the fundamental issues in limited language



usage, coupled with the belief that language interaction with skilled adults will be effective in fostering the needed organisation of experience. It is possible to argue that Tough's insights - which have been sketched all too briefly here - do in fact portray the method by which all young children develop certain language skills when interacting with adults. The crucial question is whether this same method is appropriate for children whose language development has already been seriously delayed.

A variation of this conceptual debate appears in the strong case that is made by Becker and Carnine (1978) for what is termed an instructional approach to early intervention. They consider that the failure of many disadvantaged children in the early years is a direct result of instructional failure, and that the learning rate of these children needs to be accelerated if they are to reach the achievement levels of non-disadvantaged children. They point out that the use of the Direct Instruction model in the Follow Through programme yielded higher achievements than did any of the other models. This model, known as Distar, has some similarities to the Bereiter and Engelmann approach, as it provides specially trained teachers with a highly specific set of tasks to carry out with the children.

In reporting on this work they point out that those Follow Through models which did most to teach different students in different ways, had failed. For Becker and Carnine the sequence of a teaching programme is defined by what is to be taught, not by who is to be taught. Individualisation in their programme is based only on entry level and on decisions as to when corrections are to be used and what reinforcements are to be given. They also question the concept of self-directed learning, arguing that this concept is based on Piagetian observations of children interacting with the physical environment, which clearly does teach by providing consequences in response to the children's actions. However no child can learn the arbitrary conventions of a language system unless someone in that system provides directed and systematic teaching. Likewise concepts and problem-solving skills can be more effectively taught when sequences of critical examples are carefully programmed by an adult rather than acquired through a child directed process.

The authors consider that much of the difficulty faced by disadvantaged children at school is a function of their vocabulary deficiencies, this becoming particularly apparent after several years at school. The reason for this lies in the fact that schools are designed for middle class children whose parents are capable of teaching them the language they need at home. It is the schools who fail to teach the disadvantaged children the words they need as building blocks for intelligent behaviour. Such teaching needs conscious instruction.

A slightly different approach to the same issue is described by Detheux et al (1974), who describe their use of the concept of mastery learning in a successful intervention programme with several first year primary classes in Belgium. The classes were matched with control classes at the same socio-economic level. Key aspects of the programme were continuous evaluation, both analytical and criterion-referenced, to enable the teacher to recognise difficulties and amend the teaching methods either for the class as a whole or for individual children. The goal throughout was to achieve specified levels of mastery in both reading and mathematical attainment.

The conceptual debate in regard to nursery education is discussed further by Woodhead (1976b), who reviews the whole field of intervention through nursery education in the United Kingdom and also describes a number of studies. Both the N.F.E.R.'s Pre-School Project and E.P.A. projects in three areas attempted to use the Peabody Language Development Kit, a structured American programme. What was particularly interesting about this programme was that a study by Quigley (1971) found that virtually all the participating British teachers had reservations about it, some of them being openly hostile to its aims and methods. Woodhead considers that the teachers' reservations were predictable, given their training and philosophy; they expressed concern as to whether it was right to interrupt the children's play with structured activities, especially since they believe that all the experiences comprising the Kit could be found in a good nursery. In contrast, the nursery assistants viewed the Peabody programme much more favourably.

Woodhead considers it probable that the nursery assistants were more receptive to the methods of the Kit because they lacked the depth of training in an alternative philosophy of education.

This certainly appears a valid assessment if it is accepted that the nursery philosophy should be that of minimal intervention and maximisation of the environmental richness to which the pre-school child has free and undirected access. Whether or not this should be the philosophy of nursery education for disadvantaged children, and whether nursery assistants by virtue of their (usually) humbler social backgrounds have a better instinct for what may be thought to be the prime need of the disadvantaged for structured experiences - often in contrast to the chaotic and unstructured freedom which many such children enjoy at home - is an issue which requires much fuller review and research.

Williams (1973) refers to the same theme of what should be the philosophy underlying nursery education. He points out that Elkind sees no conflict between advocacy of enrichment for the middle class child and instruction for the disadvantaged, since middle class children usually get heavy doses of instruction at home, for which the enrichment of the nursery class provides an antidote; disadvantaged children, in contrast, who experience a lack of pressure and lack

of instruction at home, need to find it in nursery school if they are to acquire the skills and motivations necessary for successful achievement in schools dominated by middle class values. Williams points out that even for Dewey intervention was clearly admissible if it increased the prospect of a child benefiting from experience; the enrichment or the improvement in the quality of the experience lay in the benefit the child derived from it.

Williams contends that in the emasculated version of the doctrine that has held sway over theory and practice in the nursery and kindergarten, 'enriched environment' has all too frequently been mistranslated into a clutter of equipment and lush provision of toys and work materials, while exegesis has passed successively through the stages of reducing to a minimum, depreciating in value, and finally prohibiting any form of adult intervention.

### 3.25 Parent intervention

What is possibly the earliest example of large scale parent educational intervention was that undertaken by the Manchester and Salford Ladies Sanitary Reform Association, set up in 1862 to disseminate 'health and knowledge' among women and children. From the outset the profession which was ultimately to become that of health visiting was conscious that it had a more integrating role than that of offering simple health advice. Owen (1977) reports that at the same time there were already in London and Aberdeen well over 100 'Bible women' who saw themselves as sanitary missionaries as well as missionaries in the more customary sense of the word. This caring movement had spread throughout the United Kingdom by the end of the century. Florence Nightingale contrasted sick nursing with health visiting: "She (the health visitor) must create a new work and a new profession for women".

In the early years of this century, and indeed in the training given up to the present time, the emphasis of health visiting has been one of dual concern, for preventive health in the narrow sense of the word and for helping mothers to foster their children's development in the wider sense, with these professionals theoretically being in a position to offer advice on all aspects of child-rearing. In the past few decades there has been pressure to integrate the health visiting profession more fully with medical General Practice and in the process the proportion of visitors' time devoted to the interests of mothers and young children has dropped from around 95 per cent to around 65 per cent. Health visiting is a uniquely British initiative, adopted by Israel in recent decades and having some limited parallels in the Scandinavian countries.

While the work of this profession, with its statutory right to call on every

mother of a young child - though without the right to insist on entry - is an example of the willingness of one service to intervene in the home, there is considerable reluctance by the much larger teaching profession to become involved in any educational activities outside the institutional framework, and equally strong reservations by some leading psychologists and sociologists about the desirability of home intervention.

Tizard, B. (1974), reporting on a series of discussions, describes the doubts of Bruner, Bernstein, the Newsoms and Joan Tough, as to the wisdom of attempting such programmes. These doubts are based partly on the belief (of Bruner and Bernstein) that the curriculum within the family is dependent on very general social attitudes and realities; Tough sees home visiting as an intrusion on the mother's privacy, while the Newsoms consider that home visiting schemes might even damage the mother's relationship with her child, if she thinks that she has been judged as inadequate. Marianne Parry considers that discussion groups in schools and clinics might be equally effective and much cheaper. Of those in favour of the idea, Ann and Alan Clarke would prefer long-term programmes focused on the mother rather than on the child, while Van der Eyken and Joyce Watt have pointed out that there are some families who can only be reached by a home visitor, whether this be a health visitor - proposed by Watt - or a woman with a less official status, proposed by Van der Eyken.

Tizard herself is critical of the American research findings in this field, arguing that the reported effects of home intervention with the child may be due to interaction with the home visitor rather than with the mother, and that what the mother learned would not necessarily be generalisable nor resistant to community influences. "Much of the thinking about parent education seems both psychologically and sociologically simplistic." Increasingly the child's learning or failure to learn is seen to be a product of the family communication system and its use, the distribution of power within the family and society, family resources - for example in regard to leisure and space - and the states of mild depression often faced by working class mothers. Changes in the mother's life may be more effective in altering her communication with her children than encouragement to play with them or read to them.

A less critical view of the potential of parent intervention programmes is offered by Blackstone (1975), based on a visit to study education and day care programmes in the United States. She cites the U.S. Homestart programme, which aims to intervene at the level of the mother-child relationship rather than simply focus on the child itself; raising the mother's confidence and aspirations and increasing the effectiveness of her achievement of these aspirations are among the goals of the Homestart innovation. Blackstone suggests that a variety of initiatives should be tried out in Britain prior to any attempt to

set up a nation-wide system of educational visitors, as suggested by Halsey in the E.P.A. literature. American experience has shown that home visitor discussions with mothers can be effective, provided they are built on the parents' own approach rather than on the 'expert' one.

An American educationist offers a further perspective on the same issue. Commenting on the problems arising from rapid changes in social structures today, Caldwell (1976) points out that despite the 'romantic mythology' which surrounds the process, parenting remains a low status job in many societies, preventing most parents from approaching the task with commitment and pleasurable anticipation. Delivery of services to parents, especially young parents, is a major task for all societies; new professional and para-professional roles may need to be created in order to accomplish this task and to ensure that the services reach the target families. At present most services offered to families needing guidance and help with child care are only offered at the point of crisis - which is generally too late. Caldwell, whose Home Inventory schedules (Caldwell et al, 1966) still stand as major contributions to an understanding of some of the crucial parent behaviours within a home environment, has also urged more research into child care environments and into the long-term effects of various forms of provision, both institutional and within different family structures, with a greater emphasis on the needs and rights of the children themselves rather than on accommodating child-care structures to the needs of adults.

While the educational intervention work of the health visiting profession in the United Kingdom has hitherto enjoyed but little research interest, the focus on the possibilities for home intervention has inevitably been concentrated on initiatives stemming from within the teaching profession. In this regard, most programmes of note are those originating in the United States. Six of these offer a contrast in goals and methods.

One of the earliest and most noted was that of Skeels (1966), who took a group of 13 orphaned children living in a mental institution and arranged for feeble-minded women inmates to care for these children on a one-to-one basis for periods of up to 18 months. Their I.Q. scores rose considerably, while those of a comparison group of 12 infants dropped; the intervention children's development was so satisfactory that it was then possible to have most of them adopted. The comparison group was left within the ordinary institutional setting, without additional stimulation. Some 30 years later all the children were traced and re-examined to evaluate the success or otherwise of their lives. As could be expected, the experimental group had nearly all made a success of their lives and were in employment; the contrast group had remained a total burden on the State. Although there were many flaws in this research, it was a remarkably bold idea. The proposal that this practice should be continued was turned down by a new

hospital administrator, some years after the success of the initial venture had been established.

The other five programmes have a number of similarities, aiming to foster cognitively useful activities through which the mothers could aid in the cognitive development of their children, though differing somewhat in their methods. Gordon (1973) offered a combined home visit and home learning centre programme for a considerable number of parent-child dyads. The home learning centres were an interesting innovation; homes were chosen from those of participant mothers and small groups of children were brought to these homes twice a week, for a few hours on each occasion. Trained non-professionals, working closely with programme professionals, were used as the intervention agents for both mothers and children. Only one of eight groups served as a control, the remaining comparisons being obtained from an interesting permutation of periods spent in the programme, these ranging from one to three years. A 'dose response' effect was noted in the various criteria used to evaluate the result of the programme.

Karnes and Zehrbach (1977) report several small-scale programmes; in one of these mothers were paid for attending weekly sessions at a pre-school centre, to undergo training in methods of stimulating their children; again some modestly useful effects were noted in the evaluation of cognitive and language skills.

Conceptually both the above programmes focus on the mother as the primary agent, although Gordon's home learning centres did direct stimulation at the child as well - through the agency of key participant mothers rather than outside professionals. Two other programmes followed a different approach. Levenstein (1977) and her colleagues emphasised modeling of verbal interaction around toys and books. What are termed 'skilled graduate toy demonstrators' were permanently assigned to particular children, visiting the homes twice a week for seven months a year. Another programme, also not focused directly on the mother, was that of Johnson (1975); the author describes how the model of  $T \rightarrow C$  (teacher to child) was specifically rejected, but so too was the model  $T \rightarrow M \rightarrow C$ , on the grounds that information provided by J. McVicker Hunt showed little effectiveness for programmes in which the only intervention was that of instructing the mother through the medium of group discussions, role-playing and other techniques. The solution adopted was that of  $T \rightarrow (M \rightarrow C)$ , with mother and child involved in the training. Both these authors report reasonably satisfactory results. Johnson's programme provided weekly visits for mothers for a year, once the children had reached the age of one year. When the children turned two the children started attending a nursery centre while the mothers attended group meetings and faced comments and suggestions from their peers, as well as being trained further.

The fifth programme of note is that described by Smith, M. (1968), in which a comprehensive programme provided the parents of 1,000 young school children

with ideas on improving the educational environment at home. As in the Hewison programme, described earlier in this chapter, parents were persuaded to collaborate with teachers in ensuring that some form of homework was done with the child each day. A particular feature of the programme was the training given to teachers in understanding the problems faced by disadvantaged parents. A comparison with children in a comparable elementary school showed a reading gain nearly double that of the intervention children.

On this side of the Atlantic there have been fewer and more circumscribed parent intervention endeavours. The E.P.A. Red House programme has already been referred to in section 3.22. Tizard (1974) reviews a small variety of parent education programmes, based in part on enticing mothers to call at what were then termed infant welfare clinics or by providing educational or other home visitors to call on mothers. A programme developed by Nicholls and Seaman at Norwich rejected the approach of showing mothers how to use toys or perform other activities according to some prescribed list, and aimed rather at helping the mothers to identify the educational opportunities in the home. It was felt that discussions about the children and their daily routines could alert parents to the children's needs, without undermining the parents' confidence. While the Norwich home visitors were trained teachers, a contrasting approach was initiated by Midwinter in a 'Home Link' programme in which mothers from the community were selected for training as volunteer home visitors. As Tizard points out, there was no built-in evaluation in most of the programmes she describes.

One parent programme which did include an evaluation component was that of Donachy (1976). Various groups were set up, some consisting of non-nursery children prior to their entry into a specific infant school, and others comprising morning or afternoon classes at a Nursery. Mothers attended the school one afternoon a week for a session with the teachers; they were given a wide choice of books to take home and offered guidance on how to use these. The mothers were also given a typed programme of activities each week, based on materials likely to be available in the homes and designed to develop vocabulary, number, time and space relationships. Control groups were drawn from a socially comparable area in a distant town. Results were rather disappointing, with only one of the four intervention groups showing significant progress in some of the indicators - mainly language.

While most of the British studies were small scale, one large-scale home intervention project is reported by Raven et al (1978) and Raven (1980). The programme took place in the Lothian region; six educational home visitors - all teachers - were appointed to the staffs of socially disadvantaged schools. Their remit was to work with two and three-year-olds in their own homes in the presence of the parents, for about an hour a week; the aim was to encourage mothers to play a more active role in promoting the development of their children.

The evaluation study found that mothers' beliefs, attitudes and expectations had changed, but not their behaviours.

The main limitation of this latter programme may have been its use of teachers to demonstrate on the children in front of their mothers. Such behaviour inevitably generates or reinforces a low self-image in the mothers, as they watch teachers achieve more with their children than the parents can do themselves. A further difficulty arises with the large number of attitudinal measures used in this study; participants in most intervention programmes usually become aware of the 'hidden curriculum' and philosophical standpoints of those administering the programmes; it becomes comparatively easy to recognise what answers are the 'right' ones when the programme researchers approach with a questionnaire on attitudes, beliefs and expectations at the end of the programme.

Some interesting variants of parent intervention programmes have been reported. Morris et al (1976) describe a relatively successful initiative in which patients at a pediatric clinic were given individual tutoring weekly at the clinic or at a nearby health centre, and asked to work with their children at home, using play materials provided by the programme. However only one-third of the parents completed the programme, out of an initial total of over 450. Van der Eyken (1982) reviews the progress made in a four-year evaluation of the Leicester Home-Start programme, in which mothers who had been referred to the programme by a variety of agencies were counselled by trained volunteer mothers. The programme was described as highly effective in helping mothers overcome their different problems, with programme organisers judging that two-thirds of families showed a marked change for the better in their morale and circumstances. Although it was hoped that the programme would also help in the development of these mothers' children, there is no clear evidence as yet that this was achieved.

Although the Playgroup movement does not regard itself as a specific form of parent intervention, attempts have been made to utilise playgroups for this purpose. The studies by Rose (1973) and Ferri and Niblett (1977) have already been cited. Both studies were aimed at assessing the effectiveness of playgroups in stimulating children's development. The results were inconclusive, offering no evidence that playgroups improved the cognitive skills of the working-class children who attended. Although it has to be recognised that a particular strength of the playgroup movement is its openness to parental involvement - that is in fact its most important characteristic - the looseness of its structure and its orientation towards middle class informality mean inevitably a reduced ability to meet the pressing need of the disadvantaged for structured experiences. Again one is faced with a clash of interests across what might be termed the socio-cultural divide.



The last major initiative meriting discussion here is the Portage programme, described in Shearer (1980) and in a considerable number of other publications. Shearer reports on a highly successful ten years of operation of Portage, with expansion to a great many countries outside the U.S.A. Essentially Portage is a well structured programme with a sequential series of target behaviours in each of a number of areas of functioning, including physical, language, cognitive and social development. Trained demonstrators teach parents how they can take over responsibility for stimulating their children's development. The key fact about Portage is that it was designed to help parents stimulate their mentally or physically handicapped children. Its claimed success in this field has led to attempts to introduce it as a technique for encouraging seriously disadvantaged parents to learn how to stimulate their 'normal' children. Portage advocates argue that the principles are of universal application.

This is certainly an appealing argument. Although there is as yet little hard research on the effectiveness of Portage compared to other forms of parent stimulation programmes, it has to be recognised that Portage is well structured, attractively designed and relatively easy to follow. Its major limitation when applied to socially disadvantaged parents is that whereas the parents of a handicapped child usually recognise that they need special and unusual skills if they are to help in the development of whatever potential their child may possess, this is not the conviction of disadvantaged parents in regard to their children. The element of rote involved in the application of Portage by parents of 'normal' children may in the end be a lot less productive than a programme geared to drawing out of the parent her own particular skills in regard to her child.

Two further initiatives are only peripherally focused on the parents. The research of Lesser (1977) and others on the success of Sesame Street has already been described in section 2.70 in the previous chapter.

An initiative which deserves mention for its comprehensiveness and broad vision of the parent and the community is that described by Pantin (1979), who reports on a highly successful extended intervention programme in Trinidad. This programme has a variety of goals, all centred on parent and community involvement and having a strong measure of democratic participation in building up the various elements within an overall development programme. Regrettably there has been no clear evaluation other than the descriptive reports of the participants. It highlights the particular problems in evaluating new initiatives in the Developing World, where not only are assessment techniques as yet unavailable or unadapted, but the whole concept of measuring results with the aid of quantifiable indicators is still seen, perhaps wrongly, as an alien Western import.

### 3.30 The methodology and research perspectives

A number of the methodological and conceptual issues which arise with intervention programmes have been referred to in the description of various programmes in the previous section. It is evident that any research is totally dependent on the quality of the methodology and the conceptual perspectives on which it is based. The clarity with which the goals are defined, the contents or nature of the intervention itself, the design of the research programme, the formation of the samples and their statistical definition, and the theoretical concepts underlying the particular programme, are all key aspects of any initiative where it is the intention to use comprehensive evaluation for judging on its effectiveness or otherwise. This section deals briefly with each of these methodological issues, prior to a discussion of evaluation itself in the next section.

### 3.31 Goal-setting and strategies

The necessity for clear goal-setting hardly needs emphasis, since a study without hypotheses of any kind is unlikely to be able to claim success for whatever findings may materialise. A considerable number of smaller studies are undertaken with little more than a vague set of objectives, such as 'helping the children to adapt to a nursery school setting', 'improving their language', or other worthy ideals. At the other extreme are the highly controlled studies in which the goals are extremely specific, being formulated in advance with a set of definite hypotheses, often with exact descriptions of the probability levels and size of effects which are sought as evidence of success.

A great deal of present-day educational research is a compromise between these two extremes, based on the fact that both the data and the methods used in educational interventions are of necessity soft or flexible. The description 'action research' is often given to the kind of study in which the goals may be fairly well defined, but the strategies are subject to considerable modification in the light of early experience with the programme.

Halsey (1972) describes the concept of action research in terms of small-scale interventions, usually within administrative systems, with a close examination of the effects of these interventions. This kind of research is particularly geared to policy issues and is not bound by the usual constraints of empirical research. Thus programmes such as Head Start, which were examples of this type of research, have been criticised because of their relatively low level of evaluation and because they were not based on a particular learning theory.

While some people argue that 'action' and 'research' cannot be linked within a single project, Halsey thinks that the interaction of the two produces a more usable and practical model for study than either in isolation. The monitoring of the research may well indicate the need to change the original research goals and parameters, especially if new relationships are discovered in a practical field situation.

Halsey considers however that a neutral evaluative role within an action research project would sacrifice the major advantage of participation in the research, since participation enables the research hypotheses to be tested in action rather than within the highly controlled situation of the laboratory experiment. He goes on to question other research procedures such as the use of a 'randomly chosen group', on the grounds that peer and other influences within a neighbourhood can lead to all kinds of problems in assembling such a group. There could also be problems in randomly allocating programmes to teachers where the teachers may be disinterested and may not have opted for participation.

The problem with the application of action research is that for many of its advocates the rejection of the unreasonable constraints of experimental laboratory research is often accompanied by a rejection of almost every firm methodological foundation on which research is customarily based. Yet this is not a necessary concomitant of action research; it should be possible to have a limited rejection of the invalidating constraints of purely experimental research and at the same time to retain the powerful insights which can be derived from high level evaluation. As pertinent examples, only a few of the E.P.A. Projects collected more than a minimum of data or subjected that data to sophisticated analyses, such as were seen in the work of Peaker (1967b), Lazar et al (1977) or more recently in Halsey et al (1980).

The question arises as to whether the undoubted quality of many of the descriptive 'action research' studies - such as the E.P.A. projects - and the profound insights offered into the issues discussed by the authors of those studies, would have been helped or hindered by a far wider data-gathering exercise and the necessary interpretive statistical analyses of that information.

In essence, action research implies a basic flexibility in the strategies and even in the goals of a research programme, with the freedom to alter these for profound reasons without necessarily abandoning the right to claim public acceptance of a set of valid findings obtained from the modified programme. Clearly the more that research principles are abandoned, the less likely it is that the final results will be accepted, at least within a research community which has the ability to judge the work. The ideal of action research should thus be to make the most frugal use of its freedom, with the retention of as much as possible of the original experimental design and the maximal use of data, in-

cluding that gathered prior to and after any changes in strategy or design.

There is little evidence of adherence to the tenets of action research in most of the studies quoted in the previous section. They tended to have fairly simple goals, with text-book designs; while these were often applied with a measure of rigidity in American studies, equivalent British studies were more flexible although this was sometimes at the expense of firmness in the interpretation of the results.

There are other debatable issues which arise when considering the goals of intervention. Should the programme itself be marginal, optimal or massive? The Milwaukee study, which has been discussed earlier, could be termed massive for its application of 7 hours intervention a day, five days a week over several years. In contrast the Ypsilanti Perry project provided what might well be regarded as an optimal level of intervention, with a moderately intensive set of programmes. The great majority of programmes, including most of the Head Start and E.P.A. studies, can only be described as marginal in that their level and period of input are minimal in relation to the need, and are equally likely to have had only marginal effects on the children.

A more intractable problem in formulating the goals of intervention research concerns the question of whether early intervention is more appropriate than later intervention. There are sound logical arguments for putting right inadequate development as early as possible, and indeed fundamental developmental research supports this approach. However the limited results of most of the intervention focused on the young child, in particular the use of pre-school institutional strategies, do not appear to offer strong support for that view. The contrasting positions of Bronfenbrenner (1974b) and Clarke and Clarke (1976) show how differently the evidence can be interpreted.

While Bronfenbrenner presents strong research evidence on the importance of early intervention, provided this occurs through the parent rather than through pre-school institutional programmes, the review by Clarke and Clarke advances evidence that critical period hypotheses have little research support, and concludes that it is never too late to attempt remediation. The Clarke's approach appears to support the views of many teachers that most corrective work can be done in the classroom - if only sufficient resources were put into this area to enable more individual work to be done with the children.

Perhaps the most fundamental issue to be decided in formulating intervention strategies is whether it is more appropriate to try to influence the child through the parent or through the institution. A growing number of research studies suggest that intervention based on parents is likely to have more abiding effects on young children than that based on institutional strategies.

While earlier sections of this chapter report on a number of studies in support of this contention, there are questioning voices. The views expressed in Weikart, Bond et al (1978) on the issue have already been cited; the team responsible for that study considered that the value of the parent intervention was largely supportive rather than definitive, in that the weekly visits by the intervention teachers to the parents' homes were seen mainly as having helped the teachers to gain better insights into the children's needs. The doubts about the desirability and efficacy of home intervention expressed in Tizard, B. (1974), reporting on the views of a number of psychologists and sociologists, as well as her own views, have also been noted earlier.

Other criticism has been voiced by Poulton and James (1975), who have strong reservations about the tenor of the Plowden Report, the National Child Development Study and other reports which place much of the responsibility for inadequate child attainment on the parents. They argue that the use made by Peaker (1967b) of (stepwise) multiple regression for the Plowden Report was dubious, since the nature of this analysis tended to highlight the influence of parent and home factors on a child's educational attainment and thereby minimised those of the teacher and school. The N.C.D.S. report in turn tended, in their view, to put the blame squarely on the home rather than on the school or on the interaction between home and school. Poulton and James see the problems to lie rather in the cultural divide in values and attitudes between school and home.

There are of course many others who stress the overriding prior importance of the home in pre-school development. For White and Watts (1973), it is the mother's direct actions and her interactions with the child in the years from 0 to 3, and especially during the second year of life, which are the most powerful formative factors in the development of the pre-school child. It is not essential that the mother should have a particular educational level or substantial economic assets. The authors are highly critical of the lack of parent education and of the research ignorance and apathy in the area of effective child-rearing practices. If anything, nursery care should emulate the practices of successful parents.

Bronfenbrenner too has been moving towards an even greater emphasis on the home environment and on what he describes as the ecology of the child; this is shown especially in the development of his thoughts over the years up to 1979. An earlier paper (1974b) reflects on the severe limitations of pre-school intervention programmes, which generally produce substantial gains as long as the programmes last but then show 'wash-out' effects in the years following the end of the intervention. While Follow Through programmes have offered the hope of prolonging the effects of pre-school intervention, one study cited by Bronfenbrenner suggests that providing the pre-school programme at a younger age (below 3) does not yield better results, if the programme duration is held constant.

The evidence does show however that highly structured curricula produce the greatest gains.

In contrast, home-based intervention has not only achieved substantial gains but these gains increase and continue to hold up rather well three to four years after intervention has been discontinued. The author cites the findings of Levenstein that the earlier and more intensely the mother and child were stimulated to engage in communication around a common activity, the greater and more enduring the gain in I.Q. achieved by the child.

Bronfenbrenner also refers to the surprising finding, by Karnes and her colleagues, that when a successful mother intervention programme was combined with a pre-school programme for the children themselves, the results were disappointing, with no difference between intervention and control groups. The conclusion of the programme directors was that the introduction of the group pre-school experience (for the children), combined with a reduction in the number of home visits, may have led the mothers to believe that they no longer played the critical role in furthering the development of their children.

In contrast to the findings in pre-school programmes that an early start does not achieve greater gains, the home programmes show that the earlier the start the better the success, with achievements being negligible when programmes start with the children already at the age of five.

Bronfenbrenner concludes that the family seems to be the most effective and economical system for fostering and sustaining the child's development; without family involvement, intervention is likely to be unsuccessful or to have short-lived effects. He also sees intervention in wider ecological terms, being needed to provide adequate health care, nutrition, housing, employment and opportunity and status for parenthood. The first three or four years of life should be focused on the establishment of reciprocal child-parent relationships, centred around activities that are challenging to the child and establish the parent as the primary agent of intervention. Pre-school intervention, in the form of a cognitively oriented programme, should follow in the last year or two before the start of school.

### 3.32 Design considerations

Statisticians usually ask that they should be consulted prior to the design stage in a research programme, since the design has a powerful influence on the nature of the analyses which can be undertaken. It is seldom, however, that advance recourse is had to statisticians, at least not in most small or medium-scale programmes. The prime design considerations are usually those of practi-

cality and acceptability by the teachers, children, parents or others who may be taking part. Frequently the demands of the specific goals and the aim of satisfying the needs of the participants leave little room for modification to facilitate more suitable forms of statistical analysis.

There is little doubt that for all research workers both the design and statistical interpretation are influenced by the conceptual assumptions or theoretical stance which they hold in regard to the issues of interest. While this 'hidden research curriculum' is an inevitable part of any intervention programme, the wider and more comprehensive the design, data collection and analyses, the less will the assumptions be likely to influence the final presentation of results - although there will always be some remaining influence from these assumptions. What all too often happens, however, is that if anything the planned design and analyses of intervention studies tend to be narrow and timid rather than over-ambitious; in such situations assumptions become of overwhelming importance - for example, if the only measures of outcomes are the scores on a particular I.Q. test, following an intervention in which a single uniform treatment was applied. With such single measures and a single block treatment there is of necessity a total conformity to the assumptions of the researcher.

Perhaps the most problematic issue in much intervention research is the univariate or bivariate approach to problems and situations which are multivariate by nature. This may be a remnant from the rigid confines of experimental design in which every possible contributor other than the treatment itself was either controlled statistically or (theoretically) randomised out of existence. The reality, as pointed out in greater detail in chapter 5 of this study, is that outside the laboratory it is impossible to have a research situation in which there are not a number of competing predictors of outcome; these need to be recognised and assessed, rather than ignored or dealt with through inadequate methods of control.

It is in the same light that designs may be criticised if they are planned so that the analyses rely entirely on an examination of bivariate relationships within separate parts of a model and within separate sub-groups of the sample; such designs are in danger of concentrating so firmly on the idiosyncracies of single trees that they ignore the causal or other ecology of the forest as a whole.

As Cronbach (1976) points out, research should have a multivariate concept of the world of research outcomes, rather than seeing this in terms of a single dependent variable; the experimenter too should treat interventions as a statistical universe within a continuous multivariate world. Response needs to be studied not as a function of a stimulus, but as a function of stimuli, innate differences, motivations, past experience, and other contributory variables.

Other aspects of design are presented by Gray and Wandersman (1980) who

present a profound and wide-ranging study of the methodology of home-based intervention studies. They point to the need to ensure, as far as possible, that the researchers' value systems and the proposed intervention changes do not ride roughshod over the interests or value systems of the target families. There is a need to take account, both in the intervention itself and in the subsequent analyses, of the part played by a variety of ecological variables such as family support systems. While the importance of sophisticated multivariate analyses is stressed by the authors, they point to the need for accompanying exploratory, qualitative and descriptive approaches to the data. The reality of what is occurring in the home between the home visitor and the mother (and/or the child) is also something needing more detailed study, as observational analyses may show a different pattern of relationships and behaviours to that reported by the visitor herself. The authors discuss the question of outcome measures; while formal instruments are of value as a criterion of performance relative to a population, broader measures of competence for both parents and children - e.g., expressions of emotional and social responsiveness, and an assessment of school success - can be valuable additional outcome criteria.

There is a great deal of useful developmental information to be obtained from properly conducted studies of this kind, according to Gray and Wandersman, provided that their longitudinal nature is recognised and the ecological environment examined together with other data. With such material it should be possible to examine the data both in hard quantitative terms and in terms of the underlying process in the development of the child.

### 3.33 Sample considerations

Even if the design of an intervention programme meets the criteria laid down to ensure a reasonable test of the research hypotheses, it is seldom that the problems of the research sample are anything but complex and on occasion damaging to the validity of the research.

While later sections of this study discuss the theoretical and practical problems which arise in the choice of intervention and control samples, it is worth noting a few of the sampling issues which have led to problems in some of the studies described in section 3.20.

As Gray and Wandersman (1980) point out, it is difficult to ensure fully comparable intervention and control groups in home intervention programmes - and indeed the problems they describe differ only in detail and severity from the problems faced in selecting pre-school or other intervention samples. A particular difficulty arises because of the recognised differences between parents who volunteer for a programme, or who agree to take part when asked, and those parents



who refuse. There are also serious ethical problems involved in denying treatment to the controls, with the possibility of the 'John Henry Effect' in which compensatory rivalry occurs if control families become aware of the existence of intervention families.

The identification of the Head Start intervention families and the recruitment, for assessment purposes, of comparable control families was one of the most difficult research problems faced in that programme. The acceptability or otherwise of the selected intervention and comparison groups was a much debated issue, both at the time of the initial assessments and with the appearance of the recent analyses (Lazar et al, 1977) of the 10 year follow-up of the intervention and control children in 12 different Head Start programmes. In recruiting Head Start families the researchers faced the dual problem of persuading parents from depressed social groups to volunteer for the participation of the children (and assessment of the parents), and at the same time to identify other parents whose characteristics were sufficiently similar for them to serve as controls - all this within a situation where funds had been granted for a Head Start initiative based on need.

The identification of the target population - and with it the research samples who would be drawn from that population - in the Educational Priority Areas programmes in the United Kingdom offered a different kind of sampling problem. The choice of the E.P.A. school areas was based on a crude indicator of the poverty of an area; schools falling within these areas were regarded as E.P.A. schools and were given additional teacher resources and special allowances for all the teachers working at those schools. Some of the many criticisms of this criterion have been cited in section 3.22.

A variety of interesting methods have been used by various research workers to overcome the basic sampling problems which arise from issues such as volunteering and the need to exclude a control sample family from the chance of participation. One of the most credible alternatives has been the setting up of a number of different intervention groups for which programmes of differing lengths are provided, with only one or two control groups. In such situations the evidence for a meaningful effect based on the length of the programme is usually quite strong evidence for the success of that programme. Other strategies involve the testing of distal control groups - groups of comparable children identified in a different area or town; however the more separated the samples areas the more difficult it may be to equate the groups.

It has to be recognised that many samples are so small and unrepresentative that research based on them may be useful only as adding limited confirmatory evidence to or raising some small doubts about the corpus of other knowledge already available on a particular technique or type of programme. Sample sizes

below 30 become difficult to interpret, while samples below 20 - for example, 12 intervention and 12 control - are so open to chance effects that only unexpectedly large treatment effects are likely to provide justification for the research. An example of this situation occurred with the Milwaukee Project (Heber and Garber, 1975), where the small size of the samples of 20 intervention and 20 control children could have imperilled the research results had it not been for the finding of mean I.Q. differences of over 30 points, after years of daily intervention.

### 3.34 Research perspectives

There are a host of crucial factors underlying the nature of the research programmes carried out in what is known as early childhood education. Three groups of authors offer particular insights into the perspectives of this form of research.

Miller and Dyer (1975), following the completion of a project in which four pre-school programmes were compared with each other, describe what they felt differentiated their project from many others. Firstly, observations were made of the intervention process, revealing the extent to which the various curricula were or were not being carried out by the teachers. The authors find it surprising that in most programmes of this kind no systematic observations have been made of this key aspect of the work; it could be regarded as a devastating criticism of much intervention research.

Unlike the programme comparisons based only on intervention/control contrasts within totally separate programmes, Miller and Dyer were able to assign children randomly to the four differing treatments and to the control group. It was also possible for the most part to control the assignment of teachers to the various curricula being investigated. Within the final sample of 171 it was possible to have replications of each of the four treatment programmes. Finally multiple measures were obtained on a longitudinal basis, enabling both immediate and long-term effects of the various pre-school curricula to be assessed; this contrasts with the practice in many programmes of focusing only on I.Q. scores.

A second study is that of Ambron (1977), who reviews the research on seven different infant and parent education programmes in a volume edited by Day and Parker (1977); she discusses nine themes common to all the intervention programmes. These include the importance of parents' influence on the personality and intellectual development of the children, the modifiability of intelligence, especially in the early years, the importance of strengthening the family - rather than trying to provide substitutes for the family in the infant years - the emphasis on language development, and the need to sustain early gains made

by the intervention. A director of one of the programmes reviewed pointed out that when someone goes into a home to teach the child, the parents begin to think there is something wrong with the way they are interacting with their own child.

Ambron goes on to note that her review of the justifications given for undertaking programmes showed up three issues. Firstly, there is a considerable gap between practice and the theory or research on which practice is based. Thus, for example, Piaget's theories are frequently cited for their importance as being one of the few which deal with the development of very young children; but what Piaget has done has been to paint broad strokes of what children can do and how they develop cognitive abilities, in the form of a natural history of child development. His theory is not a theory of intervention and does not describe a curriculum that should be followed in educating children. Likewise the provocative but isolated studies of Bernstein which pointed to the relationship between social class and restricted or elaborated codes does not necessarily provide the proper basis for a language development programme.

The 'compensatory education' mentality of the sixties is also problematical, as it motivates many intervention programmes. If views on cultural deprivation and the supposedly negative influences of the home environment are used as the basis for inviting parents to bring their children to a pre-school centre or asking them to have someone call at the home to teach the child, this can be counter-productive. Several of the studies cited show sensitivity in dealing with this matter, including those of Ira Gordon (1973, 1977).

The final issue is the question of which approach is preferable, compensatory infant or parent education, or a family support system. Ambron asks: "Why do we need to rely on a pathology model that assumes that something is wrong with parenting in this cultural group or in that income level only?" The statistics on changes in the American family, assembled by Bronfenbrenner, are staggering and justify the claim that in terms of supporting children and families, America as a whole is a backward and underdeveloped nation. While most families are in need of occasional aid in rearing their children, priority should be given to those with the greatest need.

A third study which focuses on various research perspectives is that of Smilansky (1979). He was commissioned by the World Bank to examine research on the education of disadvantaged pre-school children in developed countries, with a view to seeing whether there were lessons for the Bank which might be applied in its funding of projects in developing countries, where the emphasis would be on education, nutrition and health.

Smilansky's review indicates that pre-school, in the institutional sense, is not the 'cure-all' it was once anticipated to be. Generally children with initially lower I.Q.s show the largest gains - this may be partly a regression to

the mean - but within a few years children who are not in the special programme have made up much of the difference with the intervention groups. Very sophisticated and expensive programmes have produced longer term results, but by definition such programmes are not feasible. Studies with parents in early childhood development have been more promising, showing that I.Q. gains have been maintained for a longer time than after pre-school institutional programmes.

The study points to a number of flaws found in the evaluation of these programmes, including the use of relatively primitive designs and measurement tools, failure to obtain adequate pre and post-test data, and the paucity of numbers in the more relevant long-term studies.

Smilansky emphasises that he is basing his conclusions almost entirely on measures of cognitive development, because of their fundamental importance. (The use of this evaluation criterion is dealt with more fully in section 3.41.) The effect of Smilansky's almost exclusive focus on cognitive increments in intervention studies is a serious limitation of the review, which in other ways is a most interesting document. As an example of the effects of this limitation there is no mention of the Lazar et al studies (1977, 1978 and other related reports showing the long-term academic effects of those Head Start programmes which were properly structured).

The author does find that parent involvement programmes have produced I.Q. gains as good as or in some ways better than those of pre-school programmes, with a longer period of post-test superiority for the intervention groups. There is also evidence of a positive impact on mothers, raising their self-esteem and having positive effects on the siblings of the target children. Despite the findings of continued mother self-involvement after the termination of the programmes, with possibly increased mother-child communication and improved awareness in the mother of her role as educator and of the child's needs in infancy, the author concludes that the main limiting factor in pre-school child-oriented education is also evident in mother-oriented education, namely that real I.Q. gains do not persist.

He recommends that intervention with a potential for building the family should be considered for future programme development, but suggests that the most effective age group for such support are the adolescents.

The overall conclusions from the work of these three authors, and also from that of Chazan and Bronfenbrenner, is that the quality of much pre-school research is limited, that well-run institutional programmes do have effects on intelligence levels, and although these are seldom long-lasting there is evidence for the long-term persistence of programme effects in regard to academic attainment.

At the same time most of these authors are agreed that programmes which focus

on the parent, rather than aim at direct teaching of the child, do have long-term effects on children as well as improving the self-concept and awareness of the mothers; in this task both professionals and para-professionals appear to have a role.

### 3.40 Evaluation, cost-benefit, dissemination and policy

The design and execution of any research programme are inexorably linked to the planned evaluation of the programme and its results, and this in turn requires some form of dissemination. At a more sophisticated level cost-benefit analyses can be undertaken prior to dissemination. How the dissemination itself is handled is determined by the research goals: usually these either serve an academic purpose, with the results strengthening or challenging existing research theories or forming the basis for new theories, or they serve to influence policy and its administration. The most interesting research serves both functions.

This section deals with evaluation, cost-benefit analysis, dissemination and policy, and their relation to the intervention programmes which are the subject of the present chapter.

#### 3.41 Evaluation

Research without evaluation is an incestuous activity, since the insights it offers are only for the eyes of the participants or for those who are willing to accept the assurances of the participants about the meaning and effectiveness of what has been done.

Although it may seem unlikely that such research could have any credibility, a great many of the policy initiatives which have a research element attached to them, to monitor process and outcome, are in fact under-funded and lack the resources to carry out a proper evaluation; the final evaluation reports are shallow, tend to be impressionistic and carry little real weight. A more serious form of evaluation failure occurs when a well-funded research endeavour is set up, ostensibly to monitor an intervention programme and evaluate its results, but reports simply on the collected impressions of the researchers or of the participants themselves.

The arguments as to whether research should always have a 'public' function are highly debatable issues, influencing the nature of the evaluation and the purposes to which the latter is put. It is easy to understand claims that research should be true to itself, that it is carried out to test theories of interest to the scientific community, and that any policy implications are only a fortunate by-product of the work. Equally it is easy to accept that the public monies which fund a great deal of research carry with their acceptance an implicit obligation to offer something of benefit to the wider society. This in

turn leads to the even more complex arguments about the extent to which any fundamental research has its justification in the fact that its practical value may not be apparent at the time, but may well become so with subsequent developments. The arguments on both sides are weighty and are not open to easy resolution, for there is no clear distinction between 'pure' and 'applied' research, along what is effectively a continuum between the two extremes.

The relevance of these arguments to intervention research is more easily examined. In essence the prime goal of educational intervention research is to establish whether alterations in the pre-school or other experiences or in the environment of the child will have the effect of increasing its level of functioning above what it would have been without the intervention. This goal is closely linked with an awareness that most of the children who stand to benefit from this type of programme are living in an impoverished environment. Thus, while there are undoubtedly a variety of psychological and educational theories which can be tested as a result of the intervention, the main concern of the research is to obtain information of value to administrators and policy-makers.

What is surprising is that the prime and often the only criterion for success in such intervention programmes is the highly theoretical measure of intelligence. This is usually assessed by a conventional but rather narrow test of cognitive functioning, yielding an indicator well known as the Intelligence Quotient. Many intervention studies base the judgement of their results totally on this measure. In effect, programmes whose *raison d'être* is an improvement in the educational functioning of children from socially and economically disadvantaged environments are being assessed by a limited and rather problematic psychological measure, rather than by an assessment of educational performance.

This issue has already been referred to in the previous section, where it was seen that the judgement for or against the success of pre-school and home programmes, based on a wide-ranging World Bank review of those programmes, was determined essentially by whether or not they brought about a permanent increase in I.Q. Smilansky (1979), in the report in question, offers a powerful defence for the choice of a measure of cognitive development. He argues that cognitive capacity is the dominant recorded deficiency of disadvantaged groups; this is particularly serious because of the importance attached by modern society to increased cognitive capabilities. There is also correlational evidence suggesting an impact of low I.Q. on school failure. The difficulty with alternative measures of functioning in the non-cognitive domain - which everyone recognises as important - is that the tools available to evaluate this area are questionable and hard to interpret.

The view that 'intelligence' is the most reliable and important predictor of

school functioning is widely though not totally accepted in education and psychology. There is much less support for theories such as those of Burton White (1973, 1975, 1979), who presents findings suggesting that research should attempt to assess competence in its wider sense, a competence made up of a great many human characteristics, including personality, cognitive and meta-cognitive functioning. There is also only limited support for the practical research application of theories such as those of Bronfenbrenner (1974b, 1979, et al), whose focus is increasingly directed to the human environment surrounding the child and the interaction between child and environment.

The dominance of the cognitive criterion in judging intervention programmes has a particularly serious effect on the nature of many of these programmes. Frequently a variety of activities thought to promote the acquisition of Piagetian or other cognitive skills are introduced, when in fact the children might benefit more from a selection of pre-academic activities whose practical purpose is easier to understand and whose self-motivating nature may ensure better attainment, especially in relation to the recognised skills which older peers are seen to be acquiring in school.

It may be hypothesised that there are two principal factors underlying the emphasis placed on the importance of cognitive functioning. Firstly, instruments for measuring intelligence are among the oldest and most reliable known to psychology. The I.Q. test is for many practitioners what the stethoscope is for the doctor - and not without good reason. Secondly, the reaction against the studies by Jensen (1969) and some others who have emphasised the genetic component of intelligence, and hence its supposed immutability, has been expressed in an over-statement of the case for the environmental determination of intelligence and the assumed ease of changing this characteristic.

The Milwaukee study (Heber and Garber, 1975) did show that a drastic environmental change, persisting for a number of years, can have very large effects on I.Q. test scores. Indeed many adoption studies have made much the same point, at a much lower cost. The impracticality of the expensive Milwaukee programme for the mass of disadvantaged children does however bring one back to the problem that I.Q. scores are not easily modified, especially if the home environment remains the same. What should also be recognised is that many sample or population studies have shown that there is a wide range of academic attainment around any particular intelligence score; the predictive power of I.Q. within these situations is thus not necessarily evidence that it is the key variable to be changed in intervention.

Moreover, as Horowitz and Paden (1973) point out, whereas virtually every criterion is in some measure culture bound, I.Q. is both culture bound and culture normed. In contrast, although academic attainment is itself highly bound by



socio-cultural constraints, it is something that society treats as important in a much wider context.

It may thus be argued that since reading and mathematics are, alongside the development of socially acceptable behaviours, the most important focus of early school education, the use of these academic criteria would be less open to cultural criticism than are measures of cognitive, language or other characteristics of the child. There is as yet no clear evidence on whether an assessment of pre-reading and pre-mathematical skills in the nursery and first school years may not offer a more useful predictor of later academic performance, as well as a more valid indicator of the effects of an intervention programme, than do scores on an I.Q. test.

While the choice of indicators is one of the most important issues in evaluation, there are other issues that should be considered. Goldberg (1968) offered an early criticism of evaluation methods; there is little evidence, in most research reports, as to which elements of a programme have been effective, and which have not; nor has there been a resolution of the problem of what works for whom. Zimiles (1970), another early critic, pointed out that most of the tests used offer an index of skills and levels of knowledge, but give no evidence of the cognitive or personality processes which mediate a child's level of functioning. Programmes are geared, consciously or otherwise, towards the test content, although only a few items in most tests are related to the reality of what should be occurring in educational intervention.

Chazan (1975a), in a review of the use of tests in the evaluation of pre-school programmes, urges that more attention should be given to construct validity and not simply to the predictive validity of these measures. The author supports the view that more use should be made of criterion-referenced tests, putting more emphasis on the growth of particular skills or competencies in children and on linking tests to programme or curricular objectives. There should also be more study of the discrepancies between test-situation performance and natural-context performance, as these may give information on the child's cognitive status.

A particular difficulty with the evaluation of many intervention programmes is that the analyses are so limited that there is much room for alternative interpretations of the data. There is also at times a doctrinaire adherence to the narrowest possible interpretation of what should be reasonable caveats in regard to design, sample and statistical concepts such as normality and multicollinearity; the effect of attempts to meet these caveats fully has been that the wider validity of the work has been imperilled. Another limiting perspective has been to treat children as members of sample groups with the assumption that they share in all the cognitive, educational, motivational and other characteris-

tics of their particular group, rather than treating them (in the analysis) as individuals with their own widely differing characteristics, subject to varying levels of intervention and definable sets of home, pre-school and school environments. With the aid of more powerful and more sensitive multi-variate techniques, many programmes which have been written off as ineffective may in fact have been moderately successful in achieving their short-term goals.

One area of evaluation to which increasing attention is being given today is the process underlying the application of intervention programmes, such as the interaction between the intervenor and 'subject', the interpretation that each places on the programme goals, and the consequences for the programme of both interaction and interpretation. Unfortunately this form of study is open to abuse, as it can be a self-serving activity subject to no outside scrutiny or rebuttal. At the present stage of process methodology, analysis of events and the interpretations placed on them by participants - even when it is linked to observation or video recordings - allow a great number of competing theories to be put forward, with little resolution between them other than on the basis of their face validity. This is not to criticise the value of the insights which can be obtained from process analysis, but rather to say that it is a very new technique which has yet to achieve adequate rigour.

### 3.42 Cost-benefit analysis

The Milwaukee study (Heber and Garber, 1975) is often cited as an example of what can be achieved by the application of psychological and educational principles in a well structured intervention programme. In financial terms this is the equivalent of boasting about what heart transplant operations can achieve within a society where most people find it hard to obtain the minimum of necessary health care. The costs of offering each Milwaukee intervention child intense and often one-to-one stimulation for a full five days each week over a number of years, of training the mothers and keeping them in employment, and of meeting all the transport needs of both for the period of the intervention, are so high that the model has little relevance to what is practical even in a wealthy society such as the United States.

It is seldom that the real costs and related economic benefits of an intervention programme are assessed, even superficially. The only calculations usually made of costs are related to the research evaluation, prepared by way of an accounting submission to the funding organisation which authorised the research. The lack of research interest in the cost-effectiveness of the intervention programme itself is potentially much more serious than any neglect of accountability over research expenditure, especially in large scale interventions when Govern-

ments may be unwilling to spend further money on similar or expanded programmes. There may well be mutual puzzlement on both sides in this situation: for research teams with promising evaluation evidence on programme effectiveness the results may seem self-evident and justification for further backing; for the authorities the results may be no more than useful evidence on ways in which the service could be expanded when the present financial restrictions can be overcome. Frequently neither side has costed the long-term effectiveness of the programmes.

The histories of both the Head Start and E.P.A. programmes have much in common in their unawareness of the economic dimensions of what was being tackled. Surprisingly, cost benefit is not a new science. It has long been an integral part of agricultural and engineering research, but appears unwelcome in both the medical and social sciences. Yet in medicine today, for example in tax-financed health systems such as that of the United Kingdom, there are increasingly agonising choices to be made as to which patients should be offered expensive treatments, and even which lives are to be saved.

Of all the social sciences, education may be the most remote from the reality of accountability. The miniscule spending on educational research, as a proportion of the total education budget, stands in sharp contrast to the spending on research regarding health, social services and defence matters. It is part of the pattern of societal beliefs that education offers something that is beyond price and is its own justification. That is of course true. But in a world short of resources and facing deficit financing, the educational intervention programmes which are likely to be acceptable for further funding are those which can show tangible financial benefits for society.

One of the few, and probably the best example of competent cost-benefit analysis of an intervention programme is that presented on the Ypsilanti Perry Pre-School Project (Weber et al, 1978). It is not without significance that the three co-authors include not only Weikart but also an agricultural and resource economist, Foster. Using the concept of marginal analysis, the study shows that the internal rate of return for one year's intervention-oriented pre-schooling was over 9 per cent, with the rates for females being well above that for males; for two year's pre-schooling the rate dropped to only 3.7 per cent. The authors emphasise that economic considerations cannot be the only justification for undertaking a social investment, but that nevertheless economic factors are relevant and do justify further expenditure on this form of intervention.

It is interesting to note that the criteria used to quantify the economic effectiveness of the programme were very stringent, being limited to evidence on class retention, special education needs or institutional care. Had it been possible to quantify other indicators, such as the intervention children's increasing superiority in academic performance relative to control children in the

same classes, the effectiveness indicators may have been a good deal higher.

There are a small number of other studies in this field, such as those of Levin (1975), Becker (1975) and Becker and Chiswick (1975). The Becker studies offer a more sophisticated version of the kinds of analyses carried out by Jencks et al (1972) on other data, and present the distribution and variations in adult earnings due to different childhood factors, including in particular educational ones.

Ribich (1968), on behalf of the Brookings Institution, carried out a wide ranging study of the available information on the 'pay-off rates' of different forms of education. He concludes that vocationally oriented training exhibits a higher rate of pay-off than does general education; this is to be expected because of its direct application and the lack of guesswork as to what constitutes good or relevant education in this field. He also shows that the rates of pay-off appear to be higher from adding to expenditure in those (poorer) school districts which are now spending relatively little. However he does not find evidence that added spending on the very early or pre-school years yields better results than compensatory spending in later school years. Two compensatory education initiatives with recorded evidence on the effects of programmes at different grade levels did not clearly indicate higher rates for lower grade levels. Benefit cost ratios for five early school compensatory programmes were found to be in the region of 1.5, compared with benefit cost ratios of only 0.58 for two pre-school programmes. However Ribich reports that he experienced considerable difficulty in getting access to data on pre-school programmes, and even when he did there were only a few which were sufficiently well controlled to justify costing them.

It is possible to speculate that the limited findings of the Ribich report in regard to pre-school programmes are largely a reflection of the lack of structure and absence of clear goals in many of these programmes, so that the report does not necessarily offer conclusive evidence that structured early intervention would be less effective than similar intervention at a later age.

Rothenberg (1975) presents a perceptive methodological exposition of cost-benefit concepts, although for explanations of the actual mechanisms of cost-benefit accounting it would be necessary to consult studies such as Anderson and Settle (1977), or some of the studies referred to in Weber et al (1978). Rothenberg emphasises that any consideration of cost-benefit needs to recognise that there are competing priorities and competing population groups within a society - this is the principle of scarcity - and that the target of an economic system is to bring about the best configuration of fulfilments. Another issue is that of substitutability, namely are there alternative choices to a particular programme? An important statistical adjustment which needs to be made in all

cost-benefit analyses is known as discounting, linked to factors such as inflation and interest rates; the effect of this adjustment is that the present value of future benefit is much less than its value at the time it is realised, raising again the question of choices between strategies geared to present benefit or benefit soon to be realised - or even to current consumption - compared with what is in effect a long-term investment whose outcome is not always certain. Often the decision comes down to a crude choice between consumption and investment.

### 3.43 Dissemination and public policy

Almost invariably the results of intervention research are destined for two different groups of readers. Because the work is, or claims to be, responsible academic research, there is usually the intention to publish or distribute the results within the academic community. However the fact that a study is concerned with a public service, being aimed at improving the educational potential or even the educational attainment of the pre-school or infant school child, means that in many if not most cases the results of the study are also referred to a local or national education authority.

It may be because of this duality of focus, and the inevitable link with political policies in regard to educational and social priorities, that there has been such intense academic debate in the United States over the interpretation of the success or otherwise of a variety of intervention programmes. Earlier discussion of the Head Start programmes referred to some of the academic disputes over the results of this major initiative; there have been few academic commentators on pre-school education in the past decade who have not dealt with the issue extensively, more because of its importance for public policy than over the interpretation of the rather circumscribed sets of data on the results.

A measure of the intensity of the academic debate over more recent educational interventions is given in a series of papers in the Harvard Educational Review, commenting on the findings of Abt Associates (1977) about the Follow Through programmes - the elementary school interventions aimed at sustaining the gains made by children who took part in the Head Start programmes. Much of the debate (Becker, 1977, House et al, 1978, Anderson et al, 1978, and others) centred on the acceptability or otherwise of the test instruments and the interpretation of the large inter-site differences which were claimed to be confounding inter-programme differences.

Without becoming involved in what is a complex argument over the interpretation of data drawn from a great many Follow Through centres, with programmes differing in concept and quality and with the inevitable problems over the accep-

tability of the control groups, it does appear that some of the criticisms, particularly those of House (in House et al, *ibid*, and House, 1978) are directed more fundamentally at what are seen as the underlying political dimensions of the evaluation, at the choice of test instruments claimed to favour one or other type of programme, at the alleged failure to give parent and community groups a major say in the direction of the programmes, and at other reported shortcomings. Clearly the intermixture of conflicts over academic issues, educational policies and political goals can make it difficult to reach agreement on any findings, although in hard statistical terms the basic conclusions of Abt Associates (*ibid*) do appear to offer reasonable evidence on the long-term success of Follow Through initiatives.

Academic debate on the outcome of the E.P.A. findings in this country was much more muted, perhaps because of the cautious nature of the findings (Halsey, 1972, Smith, G., 1975, *et seq.*). In contrast, two major non-intervention educational research studies - those of Bennett (1976) and Rutter et al (1979) - which offered bold claims for their conclusions on sensitive issues, were the subject of intense debate, on both the academic and political levels.

It appears inevitable that the underlying political philosophies of researchers will be brought into issues such as the design and approach to intervention programmes, which by their nature are concerned with righting what is seen as social and educational injustice. In one way this is healthy, since it has long been recognised that philosophical concepts do influence the wider interpretations placed by academics on research findings, and indeed influence their choice of issues for research. On the other hand there are always risks in the possible influence of what might be termed a social consciousness or other strongly held views on matters such as class, intelligence, heredity and environment; this may on occasion lead to situations where academic rigour yields place to political beliefs and ideals, at great cost to the image of the neutral and honest researcher.

It is when research needs dissemination in the public forum that this temptation becomes greatest. Yet without a willingness to discuss results with a wider audience than that of ones fellow academics, research itself can become sterile. As Bronfenbrenner (1974c) expresses it, rather than claiming that social policy needs science, it is science which needs social policy in order to provide it with vitality and validity. He quotes a number of perceptive questions put to him by policy makers - non-academics - in regard to fundamental issues of child development. He was asked, for example, what evidence there was to support the policies advocated by different groups, such as half day versus full day nursery care, the mixing of social classes, the value of father care versus mother care for the young child, the desirability of age segregation at school and pre-

school, what school changes could counteract the phenomena of dropout, drugs and vandalism, the better use of television commercials aimed at children, and the design of housing developments to enhance the psychological growth of the child. He had to admit that science did not yet have clear answers to most of the problems which had been raised by the policy-makers.

The evidence of these and other questions suggests to Bronfenbrenner that much research to date has been ecologically invalid, having failed to recognise the interaction of a great many factors, human and otherwise, related to the development of the child. He considers that there is a vast amount of research to be done in these areas, and it is social policy which should provide the points of departure for the identification of significant theoretical and scientific questions.

Examining intervention research from the point of view of the policy-maker, it can be seen that the progress made by research 'technology' in the past half century has been so considerable that it is customary for Departments of State and Local Government to have research divisions of their own as well as commissioning research by universities and other academic bodies. This in turn has brought research and researchers into closer contact with political reality, leading to an uneasy but potentially profitable relationship for both researchers and policy-makers. Another link with this reality is being forged by the many charitable foundations which now sponsor a great variety of research endeavours, including work in those sensitive areas where governments may hesitate to initiate inquiry or intervention.

From the viewpoint of the researcher, the apparent unwillingness of policy-makers to accept findings and recommendations which do not match political preconceptions or fit in with other political goals is a very frustrating aspect of this relationship. The policy-maker in turn can point to the ceaseless and often bitter academic debate over the interpretation or even the validity and reliability of research findings, and can claim that such a situation demands caution before any set of recommendations can be accepted. Both researcher and policy-maker are faced by the even weightier economic arguments of those who have charge of the public purse, and who ask whether there is any financial benefit which can be expected from the proposed increase in spending - in this case on a wider intervention of the kind which the researcher claims has already brought about a set of educational and other benefits.

With the growing attention that is being given to process and the hidden curriculum, there are profound questions as to what value systems are being advocated by both the researcher and the policy-maker. How much, for example, do the proposals for an expansion of nursery education and the State funding of other pre-school facilities imply an acceptance of both the hidden and formal

curricula within these types of provision, and if they do imply acceptance then how much benefit will these curricula have for the disadvantaged 40 per cent of children? Is the current curriculum, created and directed mainly by middle class professionals, equally suited to the pre-school developmental needs of both advantaged and disadvantaged? Have those who question the universality of the value systems of researchers, policy-makers and pre-school professionals any better alternative to offer, or are their views simply a more radical middle class abandonment of even the existing structures within the different areas of pre-school provision? Does the heart of the problem lie in inadequate financing, in class bias and class assumptions, or in the failure to match provision to group needs - the failure to offer what is thought to have been inadequately offered in the home, whether it be structure for some and socialisation for others, within an overall curriculum?

The uncertainty within research itself about these issues, and the consciousness that many conceptions can be heavily influenced by the social background and education of the researcher, make it particularly difficult to interact confidently with policy-makers - with their own background and education - especially because the research uncertainty is so pervasive. However the alternative of withdrawing into ever more esoteric interpretations of ones own data, with ever fewer logical or statistical checks on whatever insights or conclusions might be claimed, may prove a worse alternative than a continuation of the public clashes over the more tangible conclusions of the research. These issues are discussed in some detail in Barker (1982).



### 3.50 The present study: concepts and hypotheses

The evidence put forward in this and the previous chapter tends to show that on many social, linguistic, cognitive and early educational indicators there are considerable differences in the nature of the attainment of the more disadvantaged and the more advantaged children in society, and that these differences have considerable effects on the school attainments of the two groups. While there is a minority opinion that it is the schools which should change their curricula and perhaps their whole philosophy to match the differing sub-cultures and attainment levels of the school entrants, majority opinion and certainly educational policy favours strategies which will improve the attainments of disadvantaged pre-school children and bring these attainments more into line with the demands of the mainstream culture which dominates education in this country. Thus, although the latter approach runs the risk of neglecting not only socially determined sub-cultures but also ethnic culture patterns, it can be argued that while education needs to take considerable account of ethnic and social minority groups, its prime responsibility is to ensure that all school-leavers have the ability to compete, as far as this is possible, within the mainstream culture and its economy.

There is also a great deal of evidence in the studies and views quoted in the previous sections that pre-school intervention among those living in situations of disadvantage is more likely to have long-term effects if it is carried out with the parents, or at least if the parents are closely involved as active agents in pre-school programmes.

The question arises of whether intervention at this level can be ethically justified. If a parent's freedom to rear a child in any way short of cruelty or gross neglect is seen to be a paramount value in a democratic society, should there not be the utmost caution about any form of parent or home intervention, no matter how mild? Such a laissez faire philosophy is in accord with the views of Herbert Spencer, who wrote of the inevitable evolutionary fate of the victims of the social system. If on the other hand it is accepted that in a modern society democracy goes hand in hand with interventionist community policies on behalf of the less advantaged individuals and groups in each community, the guidance of parents in regard to the early rearing and education of the child may be seen as ethically desirable. The question of the rights of the child is also pertinent here, since it may be argued that society has a responsibility to encourage optimal development of all children.

The intervention..../

The intervention programme devised for this study can be summarised in the following conceptual terms:

- a. It has been founded on an acceptance of the evidence that many children born into situations of social disadvantage reach school with social, linguistic, cognitive and early educational attainments different from and often well below those of many advantaged children, and that these differences contribute to lowered educational performance throughout the school years.
- b. It has been focused directly on parents rather than on their children or on pre-school institutional strategies, not only because of the evidence on the relatively greater success of parent-oriented programmes but also in the belief that most parents, given the choice, would wish to receive clear guidance on possible methods for stimulating their children's development in the years before formal school starts.
- c. For practical reasons, related to the nature of a study carried out by a single individual, teachers have not been involved in administering the programme, although they have been fully consulted about its design and content.
- d. The nature of the study and its limited resources has meant that it has only been a moderate or minimal intervention rather than optimal or massive, in the terms defined in an earlier section.
- e. The programme itself has been a semi-structured one and has been aimed at persuading parents to examine and if necessary to modify their child-educating behaviours, following discussions of the educational principles involved, rather than being aimed at modifying parents' attitudes or beliefs; while the latter is difficult to achieve, there is some evidence that behaviours found to be successful are likely to be reinforcing and self-perpetuating.
- f. The programme has been designed to test a model of intervention rather than serving as a laboratory experiment; thus, while precautions have been taken to ensure that it meets a variety of methodological and statistical criteria, it has not necessarily adhered to many of the criteria which define a laboratory experiment.
- h. The programme's ultimate purpose, outside its clear function as an integral part of the present study, is to point to methods by which pre-school parents - particularly all those living in situations of disadvantage - can routinely be offered structured guidance by school personnel on how they might prepare their children for the development of reading and mathematics in the early years at school.

It has to be accepted that there are various limitations in the study to be described in the following chapters. By its very nature and its moderate scale it has not been possible to provide the diverse and individualised programme

that one could offer a parent in her own home. There has also been concern that the programme should not be seen as prescriptive and that parents should not feel compelled to impose the suggested activities on their children if the children did not see these as enjoyable games. Much emphasis was therefore laid in the programme on the need to respect the integrity and wishes of the child in any activities that were fostered by the parents.

The broad hypotheses which were tested in the study were the following:

- A. That the intervention programmes, focused on the parents of nursery class children, will differentially influence reading or mathematics attainment in the children's first year of school according to the nature of the programme given to different groups of parents.
- B. That parent participation in the intervention programmes will prove to be a significant contributor to first year school attainment, beyond the contribution of the assessed 'parent academic environment', the child's pre-test academic skills and the child's pre-test cognitive skills.
- C. That path models of the contributors to early academic attainment will show that there are important differences between the sexes, between ethnic groups and between social levels, in the particular variables which contribute to performance.

Footnote: Although the Milwaukee project has been described as 'highly successful' it should be pointed out that criticisms of the project have been made on various grounds by e.g. Page (1972) (lack of random allocation to experimental and control groups; the tests were taught; lack of detailed description of project and its programmes.) Other criticism comes from Clarke (in Bond & Joffree (1981)) concerning possible distortion of results.

Refs.: Page, E. B. (1972) "Miracle in Milwaukee: raising the I.Q." Educational Researcher 1, 8 - 16.

Clarke, A. M. (1981) in L. A. Bond and J. M. Joffree (Eds.) Facilitating Infant and Early Childhood Development University Press of New England.

#### 4.00      I n t e r v e n t i o n      P r o j e c t

The historical and theoretical background to the present intervention project can be found in the two previous chapters.

The following section expands on the rationale for the project and describes the conceptual and methodological reasons why the intervention was tackled in one particular way, out of a number of alternatives which could have been adopted.

Section 4.20 deals with the sample. Some theoretical aspects of sampling are reviewed. The choice of sample for the project and a number of the socio-economic, employment and other characteristics of the sample area are dealt with, as well as the area's pre-school and library facilities; the selection of sample schools, parents and children, and some characteristics of the schools are discussed.

Section 4.30 is the third main section of this chapter. It presents the method used for carrying out the project. The research schedule is illustrated by way of a diagram. Individual sub-sections describe the tests used to assess the initial attainment and cognitive performance of each child at the nursery and reception class levels as well as the justification for the choice of these tests; the design of the parent interviews and the reasons for the particular method followed; the nature and detailed content of the parent reading and mathematics programmes; and finally the choice of the post-test attainment criteria.

#### 4.10 Rationale

One of the reasons for the decision to examine factors in early reading, in a study carried out by Barker (1976), was the previous finding that home background appeared to weigh far more heavily in contributing to variation in reading attainment than did some carefully assessed differences in mechanical skills reported to be highly related to reading. This is of course a well known fact.

The problems associated with measuring motivational variables, and the question of their relative importance or otherwise, as described by Barker (ibid), suggested in turn the more direct approach of an intervention programme in which an attempt would be made to influence early academic attainment among a disadvantaged pre-school sample.

As has been set out in chapter 3 of this study, the tantalising prospect of intervening at pre-school level to assist the disadvantaged pre-school child is one which has fascinated educationists and research workers for generations, starting in the previous century. While deschoolers such as Holt (1969), Lister (1971) and Illich (1974) offer a more recent prescription of seeing schools as the 'problem' and the replacement of institutional schooling by societal education and experience as the 'solution', most of the authors and intervention programmes cited in chapters 2 and 3 have focused and continue to focus on the child, the home and the pre-school institution; their belief is that one or more of these factors hold the key to reducing if not overcoming the educational backwardness which manifests itself among a significant proportion of disadvantaged children when they start school and attempt to cope with the academic demands of developing reading and mathematical skills.

The many studies reviewed in previous chapters and the varying ideological streams and prescriptions do not need to be referred to again at this point. It is sufficient to recognise that there are numerous possible approaches to the problem of pre-school intervention and that as yet there is little conclusive evidence as to the most successful and long-lasting of these methods. One of the few conclusions which enjoys fairly widespread support, at least at the theoretical level, is the concept that direct or indirect parental involvement is essential to any intervention programme, even if the programme is school-based.

It was clear that the planned intervention in the present study would need to have as its focus the pre-school institution, whether as an academically-oriented nursery class, a playgroup or other voluntary endeavour, or a child

care day nursery centre. Access to parents and pre-school children would have been difficult without this entrée, at least within terms of a limited post-graduate programme. The excellent organisation surrounding the provision of nursery schools and nursery centres in the United Kingdom (and also in the particular Metropolitan area where the research was undertaken) suggested that such institutions were likely to provide the best access and opportunity for intervention. The choice was further narrowed by the consideration that nursery schools or classes usually facilitated transfer of their older children directly into nearby or attached infant schools, so that following the children in the first year of schooling - to test the effectiveness of any pre-school intervention - would be relatively easy. On the other hand transfer from day nurseries to infant schools is more complex, the social problems of most of the day nursery children are usually more serious, and the parents are more likely to be out working than is the case with nursery class children. In the event, as is described in the next chapter, the decision to focus the intervention on nursery class children and their parents enabled a stronger research design to be formulated, once the administrative problems of gaining entry had been overcome.

Schaefer (1972) and Chazan (1975b) are among a number of writers who have reviewed the possibilities and problems of pre-school intervention and evaluation. These issues, set out in previous chapters, are intimately related to concepts about the role and academic function of the pre-school setting. Such concepts varied considerably across the schools ultimately selected in the present study, despite the apparent similarity of nursery practices in the different school settings.

Given the choice of the nursery class setting, the question arose as to the specific nature of the proposed intervention. Should it be focused on the homes of the nursery class children, and if so should be aimed specifically at the parent, the child or both? Or should it focus on the school - perhaps involving the nursery staff? The difficulties of working in the home for a male researcher had to be recognised; it would not have been easy to obtain permission to tackle a programme with that orientation. On the other hand to attempt to work through the nursery staff, even with a measure of parent involvement, would imply that the current day to day 'intervention' of nursery schooling was not achieving all that it might do; there was however no evidence to suggest that the nursery staff were not working at a highly expert level in attempting to overcome developmental lags (such as that of language) among the more disadvantaged children in their classes.

There was also the consideration of the point made by Schaefer (*ibid*), Bronfenbrenner (1974b) and many others, that the parent should be seen as the first and most important early educator of the child. Schaefer has suggested

that the restricted professional, institutional definition of education should be replaced by one in which the parent's major role is recognised. The wealth of evidence cited in chapter 2, on the links found between lack of home stimulation and poor school achievement, and similar links between high home stimulation and high achievement, suggested that it was essential to make the parents the chief focus of attention in the proposed intervention.

Clearly it would have been impossible, at this early stage of testing out an idea, to have asked that nursery staff should sacrifice nursery class time to work directly with the parents, since their own schedules, with large classes, were extremely demanding of effort and emotional commitment to the three and four-year-old children - a number of whom exhibited severe behavioural problems.

The only alternative was to organise parent meetings within the school confines, or nearly, using the experimenter as the leader of parent programme meetings in each school, and as the organiser of the details of the programmes, in liaison with early education specialists consulted by the experimenter. The experimenter saw himself in the role of a resource provider and guide for the parent groups, offering the choice of a variety of activities and ideas, while explaining their purpose, rather than prescribing a narrow set of 'formulas'. At the same time it was thought necessary to have the programmes fairly clearly structured and not merely centred around a series of topical discussions; thus a wide range of different tasks was to be suggested and parents were to be asked whether these tasks had 'worked' or not, in the weeks between successive meetings.

At all times in the development of these ideas it was the intention to focus exclusively on the parent rather than on the child, encouraging the growth of skills in the parent in the belief that in the long term a committed parent evolving her or his own strategies is more likely to remain committed after the end of the programme than would be the case if the parent was trained to carry out skills imparted by a professional demonstrating the 'right' way of educating a pre-school child. It was considered that only highly specialised skills, such as those taught by Cunningham and Spicer (1979) and others in the field of child handicap, may need to be demonstrated; even in such cases, as Cunningham (1980) has pointed out, the degree to which a parent can be encouraged to develop her own strategies rather than follow the taught strategies, is of prime importance in the success attained when working with her disabled child.

The desirability of focusing the intervention programme on the parents also became apparent when the preliminary meetings took place with the nursery staff. Nearly all of them stated that they had been asked on various occasions for advice on "how should I get my child ready for school?", or "what must I do to start him (or her) reading?" Despite the somewhat gloomy expectations that

parents might not be willing to participate in intervention programmes which required them to attend regular meetings in the nursery class building, or nearby, the staff did believe that many parents wanted at least some information on how to prepare their children for school.

While there were certain situational factors which made it fairly inevitable that the intervention programme would need to be designed and run entirely by the experimenter, with parent groups focused on the school rather than the home, the second major question of programme content was more difficult to resolve.

Language is rightly seen as being at the heart of the early education of children. Yet it is also a highly sensitive area. Debates such as those waged between Bernstein (1971), Labov (1972) and Rosen (1972), show the sensitivity of this issue to charges of cultural imperialism if language models are presented which are alien to the family's home language characteristics. Yet the greatest concern of nursery teachers is over the inadequate and limited language skills - rather than different language skills - of many disadvantaged children. Tizard et al (1980a) have pointed to shortcomings in the nursery model of fostering language development. While the issue is clearly of great importance, its resolution is uncertain and many claimed intervention approaches have aroused much ideological controversy, without necessarily achieving great progress within any one paradigm of language development. Only workers such as Blank (1972) and Blank and Solomon (1968) have presented relatively culture-free models of language intervention - but even here there is little hard evidence on the success of the particular approach followed by these workers.

It was regretfully decided that, despite the clear priority of language as a topic requiring major intervention among many disadvantaged pre-school children, the focus of the planned programme should be aimed at two other areas closely related to early school attainment, namely pre-reading and pre-mathematical skills. These skills were to be interpreted in the widest sense, but it was also the intention to include a specific academic orientation. Clearly the outdated approach of fostering mechanical skills or encouraging rote learning was rejected from the outset.

In the field of reading it was decided that parents should be encouraged to foster simple and enjoyable reading skills such as the ability to identify (and later to write) the child's own name, the ability to identify environmental words, and eventually to be able to read very simple sentences in a reader made up of a small number of illustrated themes, mostly related to the child's environment. These skills were to be developed alongside fundamental practices leading to an early interest in books and pictures, such as establishing a daily routine of reading to the child from interesting books, and where possible join-



ing a library. Other components involved games such as matching cardboard letters to similar letters written on sheets.

The programme designed around pre-mathematical skills offered an even wider range of interesting activities designed to foster early mathematical concepts. Sorting, classifying, matching and other games, using cardboard or relatively inexpensive toys and materials, were to be linked to a variety of more structured early mathematical activities such as number recognition, using special cards printed for the purpose, and games involving sets and other early concepts.

The wealth of ideas which could be presented to parents in intervention programmes meant that the problems in formulating this part of the research design were relatively minor. On the other hand the questions surrounding the evaluation came to be seen as of major importance. Chazan (1975a) has described a number of these difficulties, including the limitations imposed by small samples or by a programme focused on a single nursery school, and the use of culturally alien tests. These and other limitations, such as the facts presented in chapter 5, showing that many approaches to statistical analysis in research are simplified to the point of being misleading, or suffer from serious statistical flaws, pointed to the need for some radical innovations both in the choice of tests and in statistical methodology. The opportunity afforded by the fact that this was essentially a research study meant that it was possible to experiment more boldly than in many other formal research programmes, both in designing new tests and in developing new forms of statistical analysis in which existing flaws would at least be minimised. Thus, for example, the reading criterion used in the study (for five and six-year-olds) was a combination of three tests, a word test and a simple sentence test drawn from the literature and a specially created infant reading test created for the research study preceding the present one. Mathematical attainment on the other hand was conceptualised in two forms - as mathematical 'numeracy' and as mathematical 'concepts', the latter having Piagetian skills as a major component.

There were some serious sampling problems, described in detail later in this chapter. Again these needed to be tackled in certain ways to avoid invalidating the research. It was decided that the design would compare in particular the performance on reading and mathematics (separately) of the mathematics and reading programme groups, as well as examining performance across other subgroups. While this did not totally resolve the problem of not having an identifiable control group in a volunteer situation, it did afford a basis for evaluating the results of the programme.

From the beginning it was accepted that the design would need to be multivariate, in the widest sense. The statistical reasons for this are set out in chapter 5. These reasons are closely linked to the conceptual reality that no

performance, educational or otherwise, can be seen as the product of only a few influences on the performer; the totality of the environment surrounding each child needs to be seen as multivariate and to be assessed in those terms in as far as the evaluation instruments and the statistical analysis can cope with this reality. While there are practical limits on the degree to which reality can be taken into account, it can be highly misleading to avoid at least a reasonable level of complexity in trying to understand and present the results of any study.

One particular value of a wider perspective is that it enables a variety of speculative statements to be made about the research, and speculative hypotheses to be tested out, as urged by Kessen (1960). However difficult it may be to come close to reality in any research, the results of speculative inquiry are that much more likely to be confirmed later if the original design does not simplify the model beyond what is absolutely necessary, given the limitations of time and resources.

#### 4.20     Sample

The first section reviews some concepts about drawing a sample from a population.

The second section offers a brief description of the major social and demographic characteristics of the population from which the samples were drawn; nursery and library facilities are reviewed; the selection of sample schools and children is discussed in some detail, together with particular features of the schools themselves; finally the selection of parents for the programme groups is discussed.

#### 4.21     Sampling: concepts and limitations

Eysenck (1975) challenges the concept of a random sample, arguing that one should presume that a psychological finding applies to the whole population until or unless this is disproved, in terms of Popper's falsifiability concept. A random sample only has value if the population parameters are to be determined. His views are questioned by Ehrenberg (1976) who says that it is far more significant if it can be proved that a particular research result holds for populations which are different. This could not be learned from sampling a total heterogeneous population. The alternative to statistical sloppiness is a definitive approach towards samples.

Many workers share Ehrenberg's approach. But obtaining a random sample can be problematical. The alternative of obtaining matched samples was once popular but is now treated with some reserve. As Burroughs (1971) and Plutchik (1974) point out, the method has dangers and limitations, and the more care that is taken to match the experimental and control samples the less generalisable are the results. Questions of generalisability are also at the core of any other method of sampling.

Another difficulty arises in relation to the kind of analysis planned. A major classical approach, which has much to recommend it, is the deliberate choice of sub-samples such that one can carry out analysis of variance based on clearly defined levels of factors, with widely recognised criteria for significance. As Peaker (1967a) points out in a chapter on the sampling problems encountered in a study for the International Evaluation of Education Attainment (IEA), the very process of dividing a sample characteristic into levels to attain a more sensitive form of analysis of variance, also introduces inconsis-

tency in the meaning of the results because of the inequality in the relative sizes of sub-samples across different characteristics. Faced with this choice between sensitivity of analysis and consistency of meaning, it was decided to use instead regression analysis.

There are other problems that need to be recognised when sampling. Runkel and McGrath (1972) show that while much attention is given to randomising subjects for research, little attention is paid to randomising the environments for those subjects. Billewicz (1974) describes a number of the important ethical and methodological considerations in intervention studies, including new factors which arise during the study. He stresses the value of large samples, which enable a study of whether findings agree over sub-samples.

The need for larger rather than smaller or replicated small samples is also stressed by Tversky and Kahneman (1971). They show the faulty thinking that lies behind some of the arguments in favour of small samples; the normal requirements for significance in replicated small studies are not easily met, for reasons of probability rather than poor research. The power of the test, sampling variability and confidence intervals are all involved and make interpretation difficult. If such factors were better understood there would be more insistence on proper sized samples in the appropriate cases.

These and other studies, taken as a whole, suggested the following criteria in choosing the intervention sample:

- a. A fairly large sample would be aimed at, to allow for the inevitable attrition without endangering the size of the remaining sample.
- b. It would be a cluster sample (as defined by Burroughs, *ibid*), in which whole school groups would be used, subject to certain limitations.
- c. Rather than attempting to spread the study across a whole population, the sample would be limited to a particular area where the population could be defined as mainly disadvantaged.
- d. Criteria of randomness would be observed wherever possible in the selection of the sample components and the allocation into programme groups, subject however to ethical and methodological limitations.
- e. The research conclusions would be based chiefly on regression analyses in which individual scores rather than group scores would be used; this would help minimise the recognised limitations in the sample's randomness and in the composition of its sub-groups.

#### 4.22 Choice of sample

The research project wished to establish whether intervention programmes would succeed with the parents of nursery children in disadvantaged urban areas. It was therefore decided to ask a metropolitan authority for permission to undertake the research at a number of schools within a particular local authority area known to suffer from a high level of disadvantage. The necessary assurances were given by E as to the goals and the limitations on the kind of investigative questions that would be put to parents. The fullest co-operation was received from the metropolitan authority, its research officers and other officials.

##### 4.221 Sample population characteristics

The local area has a slowly declining population - at present about 280,000. The 1971 census showed over 20 per cent foreign born residents; today's total could be 25 p.c. (per cent), of whom more than half have come from New Commonwealth countries, mainly the West Indies. While some parts of the area are middle class, most parts are relatively disadvantaged, with occasional pockets, single streets or cul-de-sacs of wealthy families. The overall mix of the population is working class and lower middle class. The census estimates indicate 10 p.c. in the professional and managerial groups, 20 p.c. skilled and self-employed, 40 p.c. non-manual workers and 30 p.c. unskilled and semi-skilled workers.

While there are few or no areas of gross deprivation, there are a number of grossly deprived families who are entirely dependent on the social services for support and housing. Unemployment is above the average for the United Kingdom, and more than 30 per cent above that of the metropolitan area as a whole. This problem is more serious for those of immigrant descent, but the general rate is far from being the worst in the country. Apart from the troubled situation of inter-ethnic relations, whose amelioration receives much official thought and effort, there is no evidence of serious social unrest.

The local authority itself is forward-looking and has a very good record of services provided and efforts made to involve the community in planning for the future. Economically the effect of major rehousing schemes in the past has been to eliminate many of the small industries and businesses which once provided much employment. Thus two-thirds of those in employment now go into the central urban area or elsewhere for their work. Again this has a class emphasis; it is mostly the better off who go outside the area for work, while the lower paid endeavour to find work close at hand, or remain unemployed.

A considerable part of the population lives in rented accommodation. Only 20 p.c. of householders are owner-occupiers; over 30 p.c. live in local authority flats and houses, and nearly 45 p.c. in privately rented accommodation. These figures are from 1971; today's totals would reflect possibly an equal balance between local authority and privately rented accommodation. Many of the people live in flats. On the whole the quality of local authority flats is reasonable and in some cases extremely good. There are only a limited number of high rise flats.

Although 17 p.c. of all families with dependent children are single parent, compared to the metropolitan figure of 11 p.c., the proportion of single parent families with children under 5 is only 8 p.c. In keeping with the rest of the country the birth rate has shown a long decline - to nearly half the rate of 10 years ago. A national survey showed approximately 20 p.c. of mothers of children under 5 to be in full or part-time employment in 1971. The interviews carried out by E indicated that 25 p.c. of the sample mothers were in employment.

Overton and Eversley (1977) point to the combination of declining fertility and increasing employment of women as being of considerable importance for the under-fives. Although fewer children should have implied better health and education for all, in fact the slowing down of population and economic growth together has made it much more difficult to cure the mismatch of facilities.

This is undoubtedly true of the present research area, where there is a major input of services and yet a fairly high level of continuing need.

#### 4.222 Nursery facilities and the distribution of under-fives

A key factor in any discussion of mothers and children under five is the level of provision of nursery education or day care, both for the children's benefit and to enable mothers to take up employment.

Statistics for the local area show a relatively high level of provision. The metropolitan authority aims to provide nursery education for 80 p.c. of 4-year-olds and 60 p.c. of 3-year-olds by 1980. Almost half the particular local authority's planning areas are defined as priority areas for nursery provision, compared to one-third of all metropolitan areas.

In September 1976, shortly before the research project got under way, there were over 20 nursery classes attached to primary schools and several separate nursery schools. Provision of these facilities has been expanding rapidly in the area, which is now among the best served in the metropolitan region. Most of the nursery provision is part-time.

Other provision for the under-fives by that date included: more than

10 local authority day nurseries and an equal number of private day nurseries; over 70 playgroups and nearly 500 registered child-minders.

At the start of the 1976-77 academic year approximately 20 p.c. of the local authority three and four-year-olds were in nursery classes and schools, and a further 15 p.c. were in primary classes as 'rising fives'. In effect, then, 25 p.c. of the 3 and 4-year-olds not in formal school were in nursery education. Since many of the nursery classes are limited to 4-year-olds or 'rising fours', because of pressure of accommodation, possibly 45 p.c. of the local authority's population of non-school-going 4-year-olds are in nursery education.

This is an important key to a description of the sample and the population. Who are the 4-year-olds not in nursery education? Complete figures are not available, but extrapolating from a variety of data and using the rough estimate that 35 per cent of the total in each of the other forms of provision are 4-year-old children, the overall picture of the local authority's 4-year age population could be as follows:

<u>Total in age group:</u>	4,000	(figures all approx.)
In formal education (rising fives)	1,000	
In nursery education	1,300	(J)
In day nurseries (L.A. and private)	350	(K)
At playgroups	700	(L)
With registered child-minders	250	(M)
Unaccounted for (handicapped, with unregistered minders, or at home with mothers)	400	(N)

A basic question on the representativeness of the sample is whether the nursery children used in the project are representative of all 4-year-olds not yet in formal school. In other words, how similar is sample J to samples K, L, M and N? It would be hard to claim that either wealthier or poorer parents could be found making disproportionate use of any of the facilities L, M or N. Possibly the mean social class of mothers at playgroups is slightly higher than the mean for nursery class mothers, while parents at all levels make use of minders. Even the day nurseries (K sample) vary considerably in the level of social intake, although the mean intake is probably well below that of the population as a whole. Without further evidence to the contrary, it can be argued that there is no reason to believe that the research sample differed significantly from the population of all 4-year-olds in this particular local authority.

#### 4.223 Provision of books

While the local authority's educational services can be termed excellent, this cannot be said for one of the most important privately supplied educational services. Bookshops appear only within the largest shopping areas. Although there is almost no housing that is not within walking distance of at least one supermarket, most of the population are not within easy reach of a book-store. Newsagents provide the frillier forms of reading, including comics for older children, but there is an absolute dearth of quality reading for young children.

The authority's libraries do compensate to some extent for this shortfall and a number of them have a progressive policy for encouraging mother-and-child readers to join. But the libraries are not as widely spread as some of the teachers thought necessary. Moreover, there are a considerable number of parents who refuse to borrow library books (let alone visit the library) on the grounds that their children "would simply tear up the books", and because they fear the fines and criticism if any damage occurs. (In fact, it is only seldom that fines are levied when damaged children's books are returned.) It is this significant number of non-library users who suffer most from what might be termed 'book deprivation'. The theme is dealt with more fully under sections 6.20 and 6.30, on the parent interviews and programmes.

While there is a relative unavailability of books for sale, in terms of the distance to be travelled, there is a total absence of ethnic culture books for black children, other than in the authority's libraries. The quality of ethnic culture books in the libraries is good, and it can with justice be claimed that these are not often borrowed by black parents - again because of fear of loss or damage. But even in areas with a high proportion of black residents the bookshops carry few or no books in which the black child is offered a human world to which it can relate.

#### 4.224 Selection of sample schools and their children

The choice of sample schools involved an element of randomness, alongside decisions made on organisational grounds. The total 'population' of nursery education facilities was 18 established classes attached to primary schools (newly-opened classes were excluded) and several nursery schools.

In an endeavour to reach the particular population aimed at, namely, 4-year-old nursery children in mainly disadvantaged areas, it was decided to apply the following criteria:

1. Only children in nursery education would be used in the sample. The original intention to include playgroup children was abandoned for organisa-



tional and other reasons.

2. For administrative reasons it was thought preferable that children and parent groups should be centred on nursery classes within primary schools, to ensure continuity both for the children and the testing arrangements. Thus nursery schools were excluded from the sample.
3. To arrive at a fairly 'typical' nursery class child, it was decided to eliminate at the outset the small minority of children who only rarely or spasmodically attended nursery class. It would have been difficult to quantify the total attendance period of poor attenders, other than by the laborious method of counting the individual days they were present.
4. Up to the time of drawing the sample it had been the authority's policy to limit nursery class attendance (at the chosen schools) to mornings or afternoons. Thus nearly all children were half-time attenders. The few exceptions (for reasons of mother's illness, etc.) were credited with an additional two-thirds of the equivalent time - i.e. their attendance score was a total of  $1\frac{2}{3}$  times the number of half-months spent at nursery. This is explained later, in detail.

The initial random selection of schools was carried out by a research officer at the metropolitan authority's head office. It was the intention to find six suitable schools. The reasons why certain schools were excluded from the random sampling, either in advance or after the choice, are interesting because they highlight the problem of obtaining a truly random set of schools.

- a. One school known to E had previously agreed to act as a pilot school, and was automatically excluded.
- b. While one school in a relatively affluent area was accepted (when randomly chosen) to enable a contrast of children and programmes, a second school was rejected (when chosen) so as not to bias the disadvantaged sample too heavily.
- c. Two schools were rejected during sampling because they were so close to other schools already chosen that the resulting geographical pattern would not have been representative of the area as a whole.
- d. One school which was accepted was later found to be heavily involved in a language intervention programme with reception class children. This, it was felt, would unduly affect the research situation and that school was thus rejected.
- e. The head of one sampled school had a serious illness at the time the choice was made and the deputy felt that it was necessary to wait a month for his return, before deciding whether to accept the research proposals. Accordingly a new choice was made.

In total thus, out of 18 apparently suitable schools, the 'random choice' of six of them meant in effect the rejection of a further six for the reasons cited. Nevertheless, the location of the chosen schools did cover a large part of the local authority area, and five of the six schools did lie within disadvantaged sub-areas; four of these sub-areas were seriously disadvantaged and one was moderately so.

All the school heads and staff concerned welcomed the research, when approached by the metropolitan research officer, although some of the personnel expressed reservations as to whether the parents would agree to take part in the proposed intervention programme.

#### 4.225 School, class and staff characteristics

While school policies differed widely, as can be expected in a situation where curriculum is largely a matter of negotiation between education authority, head and teachers, each of the schools was committed to the major goals of language and literacy for their reception class entrants. A typical report by one school pointed to the dearth of library facilities in the area and said that the school's response was to emphasise a multiplicity of books and stories, from the earliest stage and throughout the career of every child, with the goal of adequate reading and writing by the end of the Infant years. Alongside this was the aim to develop spoken language by class discussion, improvisation and dramatic plays, children's assemblies, and the fostering of individual communication skills.

The class organisation within the schools varied considerably. In two of the schools the children who moved out of the nursery always went into the same teacher's class, until they were ready to move on into a further age-graded class. At another school the children were divided randomly into family-grouped classes; at a fourth school there was part family-grouping for the new entrants. At the two remaining schools the children's destination varied according to the term in which they moved up. Keeping track of the allocation of more than 170 nursery children, as they entered reception classes, clearly presented problems.

Of equal interest are the nature and philosophies of the nursery classes. None of them fitted any one pattern, but it could be argued that along the spectrum of policies there was one with a fairly formal approach to nursery education, placing emphasis on mild discipline and certain group activities, another with a fairly progressive approach in which creative activities were given the highest priority, and four lying within the 'moderate' middle ground. However, it was not the goal of the research to define or compare the policies

or philosophies of the nurseries in any detail. They were seen as typical of the authority's nursery classes. Adjustments that were made in the parent programmes within the different schools related mainly to administrative necessities such as facilities for meetings, rather than adaptation to any hypothetical differences in the nature of the nurseries themselves.

Virtually all the schools took a deep interest in their parents and E became aware of a variety of activities aimed to involve the parents more closely in the life and educational goals of the schools.

Without attempting to report on all those activities, some of the more unusual forms of parent liaison are listed below:

i. A regular book sale, with takings up to £20 in a session, to try to provide for parents, at cost price, children's reading books which were unobtainable in the area of one school - neither a bookshop nor library were within walking distance.

ii. A mum's club which met twice a week and was run by a Black educational liaison worker linked to one of the schools.

iii. A toy library run by a nursery class, with mothers encouraged to borrow a toy each Friday and exchange it the following Friday.

iv. A parent intervention programme (involving a small number of parents of reception children older than the sample children), which provided a small number of parents with reading and mathematical games for their children. This programme covered several schools within the metropolitan authority and was serving as a pilot project.

v. A community association, open not only to school parents but also to outsiders, in a deprived area where it was felt that the ordinary parent-teacher association would be too remote for the parents. This association ran school functions, both educational and social, and was organised on mass community lines rather than being run by a small untypical executive.

vi. Regular school meetings, two or three times a year, on purely academic topics, the aim being to show parents not only the thinking behind the school's curriculum and methods but also to encourage parents to take a more active part in their children's education at home.

vii. A regular day each week when the parents of children in one of a school's infant classes would be asked to attend a brief function where the work of the class would be displayed in visual and spoken forms.

viii. A parents' 'orange and biscuit' morning in a nursery class each week, often attended by the head; the children served their parents with refreshments.

E attended a number of these functions, usually as observer and occasion-

ally as one of the speakers.

Parents were encouraged to spend some time in the different nursery classes, although for organisational reasons it was not the policy to ask more than a few parents to be present at any one time. In only one school was this not possible, in view of the limited area of the nursery room. Here too, however, as in all the other nurseries, there was close liaison between the parents and the teacher and assistants.

The importance of this liaison cannot be overemphasised. For many parents the occasional (or sometimes frequent) discussions with the nursery personnel were the only form of regular contact with what might be termed an educational 'authority'. From hearing and over-hearing many of the discussions - usually held quite openly - it was clear that they served an invaluable purpose in both educational and social terms. Behavioural and other developmental problems would be discussed rationally and solutions sought. In theory this should be an even bigger part of nursery life, but for a teacher and assistants trying to cope with up to 30 children as well as providing a wide variety of needed activities, it was impossible to spend more than ten minutes on parent liaison at the start or end of any school session.

One particularly noticeable phenomenon was the part played by the older partner in the nursery teams. When the woman concerned was above the age of the mothers, she often served as a combined counsellor and comforter in regard to the various social crises which arose. Such crises did not occur frequently in the life of any one mother, but the availability of an older teacher or nursery assistant, to listen to the troubles and answer questions as to where help could be obtained, was seen by E as a major contribution in the particular situations of crisis. The fact that such crises occur more frequently among disadvantaged than among advantaged parents emphasises the importance of the unofficial help given.

#### 4.226 Selection of parents

The criteria by which the children were chosen for inclusion in the sample have already been outlined. As Tizard, J. (1976) points out, there are chance reasons why a parent's child attends or is admitted to a nursery class - reasons such as, for example, the degree of provision of nursery facilities within an area. Thus the sample definition of a nursery parent is already a limiting one, since not all parents who wish to send their children for nursery education can do so.

The original sample aimed at was 200 children, and their parents. This was to ensure that despite attrition losses, the remaining total would be large

enough to enable proper statistical analyses to be carried out. The details of the losses and the method of handling the differing totals for the preliminary analyses are outlined in section 6.10.

The division of parents into experimental and other groups was a difficult area, mainly for ethical reasons but also on organisational grounds. The usual aim in a research programme is to create separate experimental and control groups, to prevent contamination as far as possible, and to follow both groups until the completion of the programme. However, intervention programmes face particular difficulties because there is an unavoidable element of volunteering or agreeing to take part when invited. In the present case the setting up of a random control group would have been highly problematical, for the reasons outlined below.

The procedure for recruiting parents was as follows. An introductory letter informed the parents of E's visit and said that some kind of project of educational games and ideas for their children would be offered to them (see section 4.32). The parents were then interviewed to determine the relevant parent behaviours in regard to reading, mathematics and a few other topics. At the conclusion, the parents were asked whether they would be willing to take part in an educational programme as outlined in the letter. It was clearly essential to establish the willingness of parents to participate and attend meetings, in whatever group they were to be placed.

If, subsequently, some of the parents who had agreed to participate were then given some ostensible reason for their exclusion - "Sorry, but there are too many parents who have volunteered, and we cannot include you in a group" - this could have led to disappointment in the more motivated parents and a sense of hostility towards the intervention project among those rejected. It might also have occurred that control group mothers could have asked to take the place of experimental mothers when the latter took up employment, and again been offended when refused.

It was thus decided to accept every parent who agreed to participate, relying upon two other methods for creating comparison groups.

1. One comparison group was formed by the assignment to it of all parents who, for reasons of employment, could not take part in a programme. In many cases the parents concerned were well aware of their educational responsibilities towards their children, so that this particular control group could not be assumed to be more disadvantaged than the experimental groups. (Details in sec.6.20.)
2. Since separate reading and mathematics groups were being set up, a comparison between the reading and maths performances of the children in each programme group would provide an additional and perhaps more reliable form of comparison for both programmes.

The statistical implications of using comparison groups rather than a control group are discussed at some length in section 6.45 and in later sections.

The character of the various parent sub-groups, the differences between them, and the numbers in each group are discussed more fully under section 6.40.

The randomisation of the experimental parents into reading and maths groups offered few problems. E deliberately went visiting parents in geographically random directions, when interviewing on separate nights. There were equally random decisions that the parents visited on one evening would be asked to join one particular kind of group (reading or maths, two-weekly or six-weekly). In cases where some casting decision had to be made, coin-tossing would be used. E's only role was to ensure that a group, once set up, did grow to full size. In those cases when parents were in unfilled groups at the end of a series of interviews around one school, either the group was left below full size or otherwise the parents were re-visited and asked to participate in a two-weekly rather than six-weekly programme, for example. In only one case was it necessary to change a parent from a maths to a reading programme (there was a slightly different approach in proposing each programme). These changes all occurred before the start of the series of meetings.

Two pilot schools were used in the research, both within the same local authority area. At the one school, well known to E, all the pilot testing was done. Initially this involved 60 children in the morning and afternoon nursery classes. The reception class battery was also tried out at the pilot school, though on fewer children since most of the tests were the same as in the nursery battery. A small number of pilot children served to assist in standardising the procedure on the post-tests. Pilot school children also helped in the nursery and reception class reliability trials, with different children being used in each set.

Two parent programmes were run at another pilot school. This was a full-time nursery school. Twelve parents were invited by name, by the headmistress, to participate in the programmes. There was equal allocation into reading and maths groups. The experience gained during these programmes was extremely useful for the preparation of the final programmes. Both pilot programmes ran for eight meetings, at two-weekly intervals, and were held during the period when the main nursery testing was under way. It is interesting that dropout from these programmes was approximately the same as that from the main programmes.

4.30      Method

The formulation of the research concept, the choice of sample, the design and implementation of the project and the analysis and interpretation of the data are perhaps the cornerstones of most research programmes. The theoretical background, the rationale for the project and the choice of sample have already been dealt with.

The present section deals with the design itself, the planned intervention and the way in which the intervention was to be undertaken within the chosen sample. Although the nature of the intervention is determined by the theoretical model, it is also influenced by the characteristics of the chosen sample. In this section it is mainly the practical and to some extent the pragmatic reality which are dealt with, for no theory of intervention can specify the details in the way that the researcher's own experience, reading and sense of 'what works or is likely to work' tend to suggest such details.

In essence there are only two aspects to the design, which need to be interwoven with each other if the research is to have any real meaning. The one is the totality of evaluation techniques to assess all relevant variables within the field of research. The other is the experimental or quasi-experimental treatment itself.

There are three forms of evaluation in the present research. Firstly, early 'educational' attainment is assessed in terms of a number of criteria both at the start and end of the 22-month research phase. Cronbach (1976) has pointed out that "People are fed up with reports that summarise student attainment in a single index and then conclude that no difference between treatments can be detected". Although detailed justification for the general multivariate approach is offered in section 5.12, it can be emphasised here that these multivariate models are basic to the approach of the present research; there appears to be no other way to take account of the multiplicity of influences on human performance or of the complexity of that performance.

This same principle of assessment applies to the second form of evaluation - the measurement of a number of the cognitive variables related to the educational attainment described above. The simplistic view of 'intelligence' or 'intelligence quotients' as somehow basic and almost unchangeable characteristics of the human performer is no longer current in modern psychology, although the remnants of the decades when psychology did much to present that view of fixed human ability still survive to blot the discipline's image today.

The standpoint taken here, in line with workers such as Hunt (1961), Narrol and Bachor (1975), Vernon (1976) and Eaves et al (1977), is that cogni-

tive performance - whether described as skills, attainments or abilities - is a compound product of the environmental impact on and interaction with various characteristics of the human organism from conception onwards. There is and can be no clear dividing line between such skills and those academic attainments defined under the first form of evaluation above. All that distinguishes them is that cognitive attainment is less capable of being acquired in the short term, is more persistent and requires much greater environmental impact or interaction for major changes to occur. While it is possible to make children 'test-wise' and even to train them to succeed well on a variety of 'intelligence tests' it is not difficult to develop alternative tests of the same characteristic on which the training will have only a limited effect; whereas training in the ability to read, for example, can generalise to almost any kind of reading matter of comparable difficulty.

The distinction between the more slowly acquired characteristics such as cognitive attainment and the more readily acquired ones, such as educational attainment, yields interesting possibilities for analysis. It is commonplace that those who achieve well in cognitive attainment tests are also likely to achieve well in academic attainment tests, whereas, as shown in studies quoted by Kamii (1973), the reverse is not necessarily the case.

There are groups where the reverse situation may obtain. A large-scale cross-lagged panel analysis by Crano et al (1972) showed the expected difference between the correlations, cognitive attainment at time 1 with academic attainment at time 2 having a little larger correlation than that of 'academic 1' with 'cognitive 2'. But an inner city sample, also a large one, revealed 40 per cent of children where the reverse occurred. The authors speculated that in the inner city environment the assimilation of concrete skills may be retarded and thus such skills could be a leader in what they term 'the causation of intelligence'. This suggested to them that retardation in the acquisition of specific skills and information results in an attenuation of the rate at which higher order cognitive organisation principles are formed.

Work done previously by the present author (Barker, 1976) looked at the link between cognitive and academic attainments, in a small-scale study of reading, motivational and cognitive variables in first-year school entrants. The results pointed to a number of measures which were interrelated and appeared to have strong predictive value for reading at the end of the year. The most important of these measures was an Infant Reading Test (I.R.T.) which yields a reading score for school entrants and also correlates highly with two standard reading tests. The I.R.T. and some of the cognitive measures from that research have been incorporated in the present study.

Within both forms of evaluation, educational and cognitive, there are



problems of assessment that are almost as important as the choice of the measures or the ways in which they are defined or understood. Chazan (1975) looks at the use of tests in the evaluation of pre-school educational programmes, their difficulties of administration, the conceptual assumptions underlying global and maximal performance assessment, and even the question of whether tests have any role alongside the techniques of observational measurement of behaviours. He recognises these difficulties but considers that tests still have an important function if the testers are well trained, if the disparities between performance in test and non-test situations can be examined or at least recognised, if the construct validity of the measures is studied, and if more attention is paid to the criterion-referenced tests which emphasise programme or curricular objectives.

The third form of evaluation concerns the assessment of the parent-home environment, in so far as this is hypothesised to affect the child's early educational performance. Assessment of the behavioural environment on the basis of an interview schedule is fraught with the difficulties present in any self-report situation. A combination of observational and interview measures would have been the ideal answer, but the time limits imposed on a single researcher meant that such an extensive assessment of individual homes was not possible. Thus the assessment is based simply on the interview protocol.

The interweaving of these three groups of assessment measures and the 'treatment' programmes is not necessarily self-explanatory. As Kamii (ibid) points out, Scriven has suggested that curriculum evaluation should include three matches: that between programme objectives and the content of instruction; that between the content of instruction and the evaluation instruments used; and that between the programme objectives and the evaluation instruments. Kamii argues that possibly the most serious problem in the evaluation of early childhood programmes is the absence of the third match. For example, early childhood education programmes for disadvantaged children are judged by criteria such as I.Q. measures, on the grounds that it is only those measures which have any high predictability for academic attainment in the sixth or higher grades.

Fortunately this problem does not affect the present research. There is a highly specific match between the programme objectives - to aid in the early academic advancement of disadvantaged children - and the 'instructional content', that is, the parent reading and mathematics programmes designed to achieve this objective indirectly. There is also a match between the programmes and the evaluation instruments, although a direct link is avoided since that could yield results which are not generalisable. Finally, there is a simple link for the third match, between the programme objectives and the instruments, since it is the set of reading, mathematics and Piagetian measures

and the multivariate model incorporating these measures which will be used to determine whether the programme objectives have been attained. All the measures have been chosen for their specific contribution towards assessing the children's levels of performance and the factors thought to influence those levels.

It is the parent programmes themselves, the attempt to make parents aware of how they can influence their children's early reading or early mathematical development, that differentiates this research from an analytical study of parental influence. The introductory chapters showed that a fair amount is already known about the part played by the parent and home variables in a child's early educational attainment. The attempt to increase this influence within a disadvantaged sample is the main substance and justification of the present intervention. Such intervention is also a useful test of some of the theories in this area.

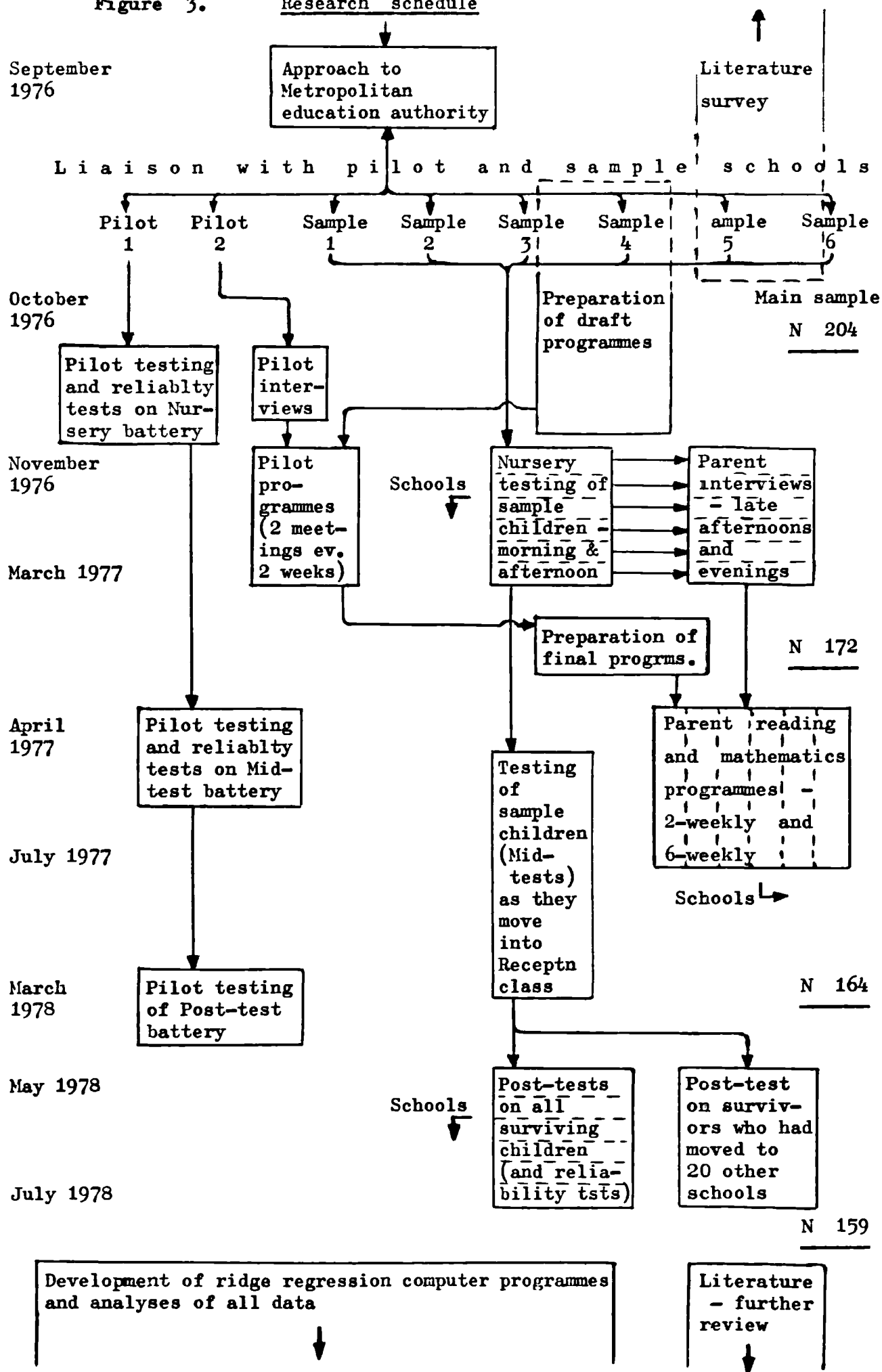
The details of the child tests, the parent interview protocol and the nature of the two parent programmes are dealt with in the following subsections.

The coordination of the different parts of the project required a fair degree of organisation in order that the work could be completed within a reasonable period. While there had to be a certain degree of flexibility in that organisation, most elements of the planned design were adhered to. The schedule itself, in its final form, is presented graphically overleaf (Figure 3.)

Further comments on the design appear in section 6.40 et seq.

Illustration ...../

Figure 3. Research schedule



#### 4.31 Child pre-tests and id-tests

In view of the importance placed within the research design on the assessment of the child's level of attainment and cognitive functioning, the methods used for assessment required careful consideration.

Many arguments could be advanced in favour of direct or natural observation. The Barker and Wright studies, launched in 1947 in a town 'Midwest' and discussed in Barker, R. (1965) and elsewhere, pioneered and confirmed new techniques and principles in this field. Social inputs were compared with behavioural outputs and theories were developed about intact 'environment force units' rather than discrete stimulus-response situations. But as Wright (1960) points out, only ten per cent of some 1400 child studies reported between 1890 and 1960 relied on methods which left nature and society to their own devices. Brandt (1972) and Willems and Rausch (1969) review the differences between naturalistic observation and experimental (test) situations and describe the advantages and limitations of the former. As Barker (ibid) points out, the experimentalist has difficulty in explaining the complex problems arising from environmental factors such as ghettos, poverty or population density, and tends to rely on the 'black box' approach. Workers such as Hutt and Hutt (1970) have shown the effectiveness of this method for child study.

However, the natural approach would have had major limitations within the present study. The criterion of academic attainment relies on the contributions and influence of a considerable number of factors; to assess each of these through natural methods would have required large amounts of time with a small and possibly unrepresentative sample, observing behaviours which in many cases appear naturally only at long and irregular intervals. Despite the naturalness of the setting and observations (for example, if children's behaviour is observed over long periods in the classroom, and related to their academic performance) there are possibly as many methodological problems with the naturalistic method as there are with the experimental approach. Johnson and Bolstad (1973) review the naturalistic problems in detail.

Accepting that the need to quantify a large number of variables, on a reasonably sized sample, demands the experimental or test approach, what are the limitations inherent in this? Cronbach (1961) shows how the degree of structure, motivational factors and the generalisability of any particular test are all relevant in the testing situation. In a series of articles brought together by Badia et al (1970) the different authors point to the difficulties which arise from the social demand characteristics of the experimental situation, the influence of experimenter characteristics, and the many fallacies underlying the 'proof' or conclusions drawn from the use of significance tests to assess experimental effects.

Reference has been made earlier to the discussion by Chazan (1975a) of the evaluation of pre-school research, the methods that are employed and the usefulness or otherwise of tests in the pre-school educational programmes; there may be a need for new tests that are criterion-referenced rather than norm-referenced. As Sigel et al (1973) point out, reliance on I.Q. scores and traditional psychometric measures often does not allow for the testing of specific behavioural outcomes; for their child intervention study these authors prepared a battery of new criterion tests.

There is a further problem area which has up to now largely defied efforts to quantify it in any meaningful way. It is known that responses in test situations are often related as much to a child's problem-solving strategies as to its cognitive skills. The development by Kagan (1965) of his test of cognitive style is one of a number of efforts to derive a statistic in this area. Banta (1970) devised a number of tests to measure 'autonomy' in problem-solving. There have been many idiosyncratic tests of the characteristic of distractibility, which is often a crucial factor in both the test situation and the general classroom performance of some children. Other areas which have only been examined with highly task-specific instruments are characteristics such as perseverance, apathy, response perseveration (as an alternative to solving the new problems), playfulness and recklessness.

A further crucial problem is posed by Sigel (1974) who asks, "when do we know what a child knows?"; he shows how the many facets of context, rapport, the meaningfulness or otherwise of the tasks, attentional and other factors can influence the child's ability to reveal what it knows.

Given these limitations - limitations which it may well be argued are often present in the classroom situation and affect academic performance as much if not more than they do test performance - and given the related problems of reliability and validity, which are dealt with in sections 5.13 and 5.14, it was decided to rely where possible on suitable existing instruments, but also to develop new ones or make use of tests developed and assessed by the author in a previous research study (Barker, W., 1976). Most of these tests were thus not standardised or norm-referenced; even in cases where standardised tests were used, such as the four sub-tests of the WPPSI battery, the scores were employed in their raw form and subjected to a different method of age correction. Clearly such treatment of test scores affords no opportunity for comparison with population norms.

However, two of the four criterion tests which are central to the research are well-known standardised reading tests, scored as described in 4.34; a third reading test, developed by the author for pre-school and first-year school children, has shown high correlations (in earlier research) with the two standardised

reading tests; the fourth criterion is a combination mathematics test consisting of the whole of the WPPSI arithmetic test, a number of the simpler items taken from the Boehm test of basic concepts, and a small number of additional items to stretch the 'ceiling' of this test. The remaining tests, administered at the beginning and mid-way through the programme, help to identify attainments and the cognitive skills which contribute to the final criterial performance and thus enable a more accurate assessment of the effects of the intervention programme.

In general the aim was to utilise short-range, relatively simple tests rather than lengthy compound tests which embrace a range of variables. Although test instructions were formalised and adhered to by and large, they were adapted where necessary to aid the understanding of slower children, or abbreviated in the case of brighter children who tended to become bored by the repetition of formal test protocols.

Considerable effort went into the establishment of rapport with the children. E spent a full day in each nursery class, at the pilot and six research schools, prior to the start of testing. Even after that a start was made only with testing the bolder children in the class, leaving the more reticent children until the last. Tests were carried out in a corner of each nursery, in full view of the other children but with the infant test table and low chairs shielded by movable furniture. Reception class children were usually tested in a relatively quiet area, away from the classroom itself, but this posed little difficulty as E was by then well known to the children.

The physical positioning of the test child and E was also determined by the need for a high degree of rapport. Rather than E sitting in a somewhat threatening position on the opposite side of the table, he was seated at one of the table corners nearest to the child, thus facing the testing instruments rather than the child, who was watched only from the side. A single case study reported by Davidson (1977), following a spell as a medical student worker in an impoverished area of Tunis, suggests the importance of this positioning for small children. Among a group of mothers given nutritional supplements and also clear demonstrations of how to cook the food, there was one seemingly concerned mother whose infant made no progress and remained severely malnourished. The mother was asked to cook a sample meal, and did so successfully. Workers then accompanied her to her home and watched her serving food to the child. The other mothers normally sat next to their child, facing the food, but this mother sat on the far side of a small table, trying to feed the unwilling child. She was asked to move next to the child, and the child immediately responded; the problem was apparently at an end. The fact that psychological test instruments for young children are generally numbered on the assumption that the E is sitting opposite the child, suggest that the example of this solitary mother is fairly general in the testing situation.

One further aspect of the test administration was the use of incentives at each stage of testing. In the nursery stage the children were given a choice of a tiny packet of peanuts or sultanas at the end of the session. With the first child proudly bearing her 'prize' to show the teacher, it was soon realised that the test experience was accompanied by the award of a gift, which E presented 'for all the work you have done'. In the reception class mid-test and post-test, a fibre pen or other small gift was offered. Drabman and Hammer (1977) reviewed the use of incentives in early childhood education programmes and found a wide variety of material, token, social, verbal and activity incentives. They discuss the arguments that extrinsic reinforcements depress intrinsic motivation but find the case not proven.

Prior to the start of testing in the six research schools the tests were administered to an entire class of nursery children at a pilot school. This enabled the procedures to be tightened and altered where necessary, and also led to a drastic change in the form of one test, that used for distractibility; two other tests, one which assessed behavioural attitudes in relation to the rights of others, and the other which determined self-concept, appeared to work selectively and after further use at a few of sample schools they were discarded. The mid-tests and post-tests were also administered first at the pilot school and a major part of the reliability assessments was carried out there (using different children for each battery of test-retests).

The children's state of needs, in terms of the first two items in the Maslow hierarchy of needs, were assessed by the nursery and reception class teachers. Details of this are also given below.

In the discussion of the individual tests which follows, the rationale for the inclusion of the tests is reviewed.

#### 4.311 Pre-tests (nursery)

The purpose of these tests was twofold: firstly to establish, as far as possible with this age range, the level of attainment in pre-reading, pre-mathematics and verbal competence and the level of awareness of what reading and books are about; and secondly to assess cognitive functioning as defined by a battery of tests on characteristics thought to relate to later reading and mathematical performance. For clarity of analysis it was necessary to make a formal division between 'level of attainment' and 'abilities', although there is much conceptual overlap. For example, a change in I.Q. 'ability' is often cited as evidence of the effectiveness of intervention. But since the present research aimed at specific improvement in academic attainments it was felt necessary to differentiate between characteristics which are closely related to the attainment concept and those 'ability' characteristics which are supportive in a more general sense.

The protocol for each test is described in full in the Appendices A.

#### Attainment tests

1. English Picture Vocabulary Test. A research project by Bougere (1969) on factors related to later reading performance showed that the addition of language measures added considerably to the variance accounted for by a reading readiness test. Moseley and Moseley (1976) bring together some of the evidence relating linguistic competence to reading and note the difficulty of defining the language that children know and the language that they need, while Gulliford (1969) points to the large number of pupils within the average range of ability whose progress is retarded by language inadequacies.

A complete assessment of linguistic ability would require the use of composite tests such as the Illinois Test of Psycholinguistic Abilities or other instruments which require much time to administer. Even the I.T.P.A. itself would present difficulties of interpretation. Pumfrey (1976) reviews several studies in which performance on this test has been compared with reading attainment; he found that different measures on the I.T.P.A. profile showed differing levels of correlation with reading. Rosenfield (1975) found a non-significant correlation between the I.T.P.A. and reading ability five months later, but significance for some of the sub-tests; moreover, no specific weakness in an I.T.P.A. sub-test was related to reading performance if it was compensated for by a strength in any other area of this test.

In the circumstances it was thought preferable to rely rather on the English Picture Vocabulary Test (Brimer and Dunn, 1973) which offers a fair indication of verbal comprehension and may thus be regarded as important to early reading per-



formance, where comprehension of language plays a larger part than does its expression. One study supporting this hypothesis was that of Francis (1974) whose research indicated that differences in reading ability between children within a social class group were related to vocabulary but not to speech structuring skill.

The widespread use of the E.P.V.T. and its American equivalent the Peabody Picture Vocabulary Test suggests that this kind of test is generally accepted as a measure of listening vocabulary. Lovell (1972), in a Buros review, points out that its high correlations with other vocabulary tests suggest that it does measure some common vocabulary function. But he argues that a difficult question arises regarding the validity of the E.P.V.T. in relation to reading problems and other verbal handicaps.

The fact that the E.P.V.T. also correlates .8 with an intelligence measure such as the Stanford-Binet (for a normal sample, quoted by Brimer and Dunn, *ibid*) is to some degree an unavoidable complication for the analysis, since both the E.P.V.T. and the four WPPSI subtests used here are likely to contaminate each other.

2. Reading awareness. This test, consisting of a short series of questions to the child about its interests, what are books for, what are words for, etc., was found in previous research by Barker (1976) to correlate .37 with a combined index of reading performance nine months later.

Studies by Reid (1966) and Downing (1969) have pointed to the importance of a child's concepts of what reading is about; Reid's findings emphasise how reading is seen as a 'mysterious' and thus difficult process. Both authors concentrate on the children's ideas of 'word' and 'sound'; this seems rather advanced for their five-year-old samples, whether or not phonic instruction was being introduced to the children at that age. The test used in the present research asked only one question about the significance of 'words' but went into broader issues about books and reading for the remaining questions.

3. Infant Reading Test. Pumfrey (1977) offers a wide ranging discussion of the rationale behind the measurement of reading abilities, the selection of tests and the problems of administration. This is a companion volume to his 1976 synopsis of the tests themselves. He cites one study by Lennon which used factor analysis to show that four major components of reading ability can be recognised and measured reliably. These are: a general verbal factor, concerning the extent of the child's vocabulary; the comprehension of explicitly expressed content, which involves reading for the literal meaning of a written passage; the comprehension of implicit meaning, which measures more advanced reasoning skills; and appreciation of the author's intentions, feelings and thoughts. Pumfrey also reviews the rather confusing range of tests which attempt to identify

and measure sub-skills in the reading process; these tests offer some guidance but lack a strong theoretical basis.

Pumfrey does not consider these two approaches to be incompatible and the present research makes use of insights from both. While the ultimate reading criterion for the intervention programme is the level of competence as established by a formal academic reading test, the assessment of initial skills covers the first two of Lennon's factors - the E.P.V.T. for verbal comprehension, and an early reading test - as well as a small range of hypothesised contributory skills, described later in this section.

The problem of devising or finding a suitable test to assess children's state of pre-reading or early reading was not easily solved when it was tackled in earlier research by Barker (ibid). Reading readiness tests are often used in this situation as a measure of children's potential prior to classroom instruction. Thackray (1965) has reported research showing correlations from .42 to .59 between six-year-old reading and earlier performances on the Harrison-Stroud test or other indicators of readiness. More recently Downing and Thackray (1971) presented a reading readiness inventory devised for English children. In the United States the use of such tests is widespread.

But readiness tests have also been widely criticised. Durkin (1976) considers these tests unnecessary and indefensible, particularly as they make use of composite scores covering a variety of skills and fail to recognise that different kinds of reading instruction require different abilities. Cochrane (1976) cites a Canadian study which found kindergarten teachers' predictions (in a sample of 75 teachers) to be superior to that of five readiness tests. Farr (1972) and Smith (1972), in Burow's reviews of one particular readiness test, point to the ineffectiveness of these tests in identifying early reading problems; even as a predictive instrument the test in question achieved only the same level of reading prediction as that of most group intelligence tests. As Durkin puts it, quoting Ausubel, readiness is 'the adequacy of existing capacity in relation to the demands of a given task'; these demands are closely linked to genetic, maturational, experiential and learning factors.

Given that some reliable indicator of early reading or pre-reading competence was seen as essential to the research into early reading attainment, a test was devised by Barker (ibid) for his earlier research project. Termed the Infant Reading Test, it assesses the child's abilities on a range of simple reading skills, starting with the ability to write one's name and to write a few other words of own choice, knowledge of the letter names (lower case and capitals), and ability to identify a number of environmental words in different fields - thus there are sets of shop words, food words, street words, television words and 'school words'. A child able to read half of any set of ten words (apart from

the last-named group) is awarded full marks for that set. The choice of words is determined by the area in which the sample children live and by the kind of words that an average child is likely to meet with at intervals within the area. A new element added to the test for the present research is a set of five questions taken from Clay's Sand Test (1972) in which the awareness of directionality in reading is assessed.

The performance of this test in the author's earlier research was satisfactory. A concurrent test of the I.R.T. correlated .89 with the Southgate Reading Test and .62 with the Daniels and Diack Sentence Reading Test on a sample of 32 inner city children at the end of the reception year. Meanwhile the I.R.T. administered at the start of the reception year correlated .72 with Southgate nine months later and .71 with Daniels and Diack. These latter correlations are well above the range of figures presented for most readiness test correlations with reading at the end of first year in school. There seemed to be no reason why the I.R.T. should not serve as one of the main instruments in the present research, with Southgate and Daniels and Diack also forming part of the final post-tests.

4. Mathematics test. While there are various projects for the development of early mathematical concepts (e.g. Matthews and Matthews, 1974-1977), there is much variation in the approach of different authors towards mathematical competence. For some, typified by the work of Clarke and Mason (1969) and the Birmingham E.P.A. project (Lines and Widlake, 1971), mathematical development is linked closely with the growth of Piagetian skills. Stahl (1973), on the other hand, reports a study that found no important differences between groups (in first grade) formed on the basis of ability in one or more Piagetian tasks, and later computational skills such as addition or subtraction.

For older children it is possible to define mathematical competence in wider terms and to examine performance in different areas. For younger children and school entrants the range of performance is more limited and computational skills inevitably play a large part in defining mathematical competence. Typical of this latter approach is the study by Swenson (1973) on teaching mathematics to children.

A more profound mathematical rather than computational view of young children's development is put forward by authors such as Lovell (1971a, 1971b), Sinclair (1971) and Soviet authors such as Krutelskii (1976). Although Lovell pays tribute to the deeper understanding given by Piagetian theories, his presentation of the growth of mathematical understanding in the early years from kindergarten onwards is by no means narrowly tied to those theories; he shows how even at young ages it is possible to foster important basic concepts. Krutelskii's major study argues that the concept of 'mathematical abilities' when applied to

younger children is somewhat provisional, although individual components of such abilities are formed even in the primary grades. He distinguishes between two of these parameters - the formalised perception of mathematical material, in concrete terms by the less able and in relational terms by the more able, and the generalisation of mathematical material, developing from something imposed by external (teacher) necessity towards something evoked by internal need.

Despite the conceptually appealing nature of these studies, it was felt that given the limitations of time and the fact that it was the academic ability of five and six-year-olds that should be tested as the criterion of mathematical performance, the only test possible in the circumstances was one which would assess the understanding of a few basic mathematical terms and measure performance in simple computational tasks. At the same time it was decided that the parent mathematical programmes would be more clearly oriented towards the development of concepts in the wider sense described by Lovell.

The test assembled for this research consists of three parts. The first is a range of ten pictures portraying relational and other concepts (biggest, nearest, under, rectangle, etc.), mostly drawn from the Boehm Test of Basic Concepts (1970). The second part is the WPPSI arithmetic test, and the third part is a group of five calculations slightly more advanced than those given in the WPPSI test, to provide a higher ceiling.

One conceptual limitation of this composite test is that the WPPSI arithmetic sub-test forms part of the WPPSI verbal rather than performance battery. However the WPPSI manual (Wechsler, 1967) presents data showing equal correlations of .56 between the arithmetic test and both the verbal and performance batteries, at the age of 4. Even at the age of  $5\frac{1}{2}$  (possibly the median age for the post-test of mathematical ability in the present research), the correlations are .71 with the verbal battery and .66 with the performance battery. It is only with increasing age that the divorce between arithmetic skills becomes wider; thus Lohnes (1966) shows that for adolescents an arithmetic test loaded highly on a factor of English and verbal skills, rather than on a mathematical factor.

In the absence of any clearly accepted criterion for early mathematical competence (which would in any case need to be seen as multi-dimensional), and given the need for a concise test suitable for four and five-year-olds, the compromise choice of a mixed test of concepts and arithmetic attainment was seen as the only viable possibility, while recognising the limitations of this instrument.

## Cognitive skills

5. Rhythmic tapping. Auditory perception has often been studied in terms of its relationship to reading ability (Vernon, 1957, is an early example). But generally the contrast has been made between 'normal' readers and the special group of backward readers who face major difficulties. Such research isolates an extreme group for purposes of comparison. In view of the evidence that backward readers are not part of the normal continuum of readers (Yule, 1974) this approach may conceal the relationship of perceptual skills to reading within the range of ordinary readers. Another problem is that many studies present evidence on concurrent correlation between reading ability and perceptual skills; this is clearly open to even greater misunderstanding than are longitudinal correlations, as the chance of fortuitous third factor causation is particularly strong in a cross-sectional relationship.

There are various forms in which auditory perception has been studied. Groff (1975) reviews a number of studies in the area of auditory discrimination and questions the current belief that there is a close relationship between such discrimination and reading achievement; in particular he cites Frank Smith's explanation of why only minimal use can be made of the complicated system of spelling-sound correspondences in English, when learning to read. A different interpretation is that of Bakker (1972) who develops the concept of temporal order of perception and presents much research on its link with reading. A theoretical language model is that of Lashley (1967) who considers that speech shows many examples of integration between spatial coordinate systems and temporal or rhythmic sequence systems; it is not difficult to link his theoretical speech model to a hypothetical reading model in which the rhythmic sequence systems interpret visual input, linking his model with that of Bakker.

Evidence from such studies suggested the use in the present research of a simple test of rhythmic sequencing ability. In the author's recent research a test of rhythmic tapping correlated .44 with a composite reading measure nine months later, while De Hirsch et al (1966) found a correlation of .30 between an auditory tapping test and reading two years later; according to Bakker however there may be a strong sex factor also present in such a relationship, with girls showing a high correlation and boys a low one. The sequences in the test used here were based largely on the auditory-visual sequences employed in the Birch and Belmont (1964) integration test, though in the present case the child was asked only to imitate a tapped auditory sequence. A preliminary study in the pilot sample, in which all the items were administered (instead of stopping after three failures), confirmed the order of difficulty, with one minor change.

There are two criticisms of this test which may have limited its effectiveness. Since the taps were made by E in front of the child, a child with a

skilled visual memory could have made use of that skill when copying the simpler sequences rather than relying on auditory memory for the rhythms. Moreover, as the child's gross motor skill is also involved in this task, it was occasionally necessary to give children the benefit of the doubt when there was a case of relatively poor manual control (that is, in judging whether the sequence was copied correctly or not).

6. Matching familiar figures. A fair number of tests have been developed to assess the personality variables which influence cognitive performance - otherwise known as 'cognitive style'. One of the best known of these is the matching familiar figures test (Kagan, 1965). Other measures include tests of attention span, pictorial distractibility, field dependence, and conceptual styles. Denney (1974) assessed some of them and found a rather uncertain relationship with reading ability. Further areas of importance concern characteristics such as anxiety and sense of autonomy, but here too the research, though widely pursued especially in the field of anxiety, reveals the same uncertainties as do most of the other tests of cognitive style; they show only a rather confused relationship with reading. The problem with almost any assessment of personality variables is that the tests are nearly always closely bound to context.

Apart from the subjective assessment of distractibility, described later, the only cognitive style test which it was felt could stand up to use in the present research was the M.F.F. The characteristic of reflectivity/impulsivity is seen to be important by many teachers, and the M.F.F. does appear to provide some evidence for this characteristic - given the limitations of context. However, the test is itself a seriously flawed instrument. Block et al (1975) showed that claims for the correlation of error, latency (time taken to make a choice) and interaction, among young children, do not hold; response accuracy (error) is the important factor for young children and was the primary discriminating variable in their research, regardless of age.

Ault et al (1976) have pointed to the low reliability of errors on test-retest and the low internal consistency for the error items; they point to the major problems caused by the attempt to dichotomise both the error and latency scores, giving quartile samples rather than relying on multiple regression techniques which would use all the individual variation. Many of the studies which have adduced that there is no significant relationship between the dichotomised groups and reading have been based on this crude division - a division which is at odds with the concept that almost any human characteristic ranges over a continuum, usually with a normal distribution. Perhaps the most serious criticism is voiced by Achenbach and Weisz (1975) who found that measures of mental age at time 1 were better predictors of the impulsivity/reflectivity quartile at time 2 than was the original I/R measure at time 1. The authors consider that the

I/R measure may primarily reflect differences in development rather than in basic cognitive style. However, their research findings were also based on the same limited dichotomisation that has been previously referred to.

For technical reasons a specially devised set of pictures was used for this test, rather than the Kagan M.F.F. or its junior KRISP form. The KRISP form correctly offers the young child a choice of only four pictures (rather than six in the ordinary M.F.F.), but these are grouped below the target picture; this may well affect the search strategies of children, and many could opt for picking one of the two pictures closest to the target. In the test used here the four pictures are spread around the target picture. The pictures themselves (reproduced in Appendices A) were drawn with a view to the age levels of the children using them - mainly four and five-year-olds - on the principle that it should be possible for virtually any child to pick the correct picture if it had the patience to study the choices. In the light of the evidence of Block et al and others, only errors were measured. (It can in any case be hypothesised that latency is likely to be more strongly related to a child's cognitive level than are errors; thus paired bright and slow children, with the same theoretical levels of reflectivity, might not differ much in their error rates but could differ widely in the latencies or speeds with which they are able to reach their conclusions.)

For the nursery battery a combined visual-visual and tactile-visual M.F.F. test was used, but in the assessment of reliability it was found that the set of tactile items had virtually negligible test-retest reliability, and thus only the visual-visual part of the test was retained. For the reception battery the visual-visual portion of the test was extended to twice the original length.

7. Bender Gestalt. In the area of visual perception the Bender Gestalt test (Koppitz, 1960, 1975) is a highly regarded instrument. It is strictly speaking a test of visuo-motor integration and is described as such by Koppitz, in preference to claims that it represents visuo-motor functioning. In a Burows review Kitay (1972) sees the test as offering unique contributions to the evaluation of perceptual-motor functioning and to a number of emotional problems. The test has been more generally used for the latter purpose and is seen as an important diagnostic instrument for children with serious reading and other difficulties. Koppitz herself, in her 1975 review of a great number of studies on this instrument, points out that interpretation of the Bender Gestalt depends on the age of the child. She considers that the use of the test on immature children below five is complicated because of missing or incomplete diagrams, or the lack of scorable details; however she does report the use of the instrument on three and four-year-old children. Koppitz questions whether the test is at all related to reading, although both the test and reading require a minimal level of maturity in

visuo-motor perception. Whereas the Bender Gestalt discriminates between children with reading problems, when the children are divided into four groups according to their Primary Mental Abilities scores the test does not discriminate between normal high and low readers in three out of the four groups.

This is an important comment, but again many of the studies on which Koppitz relies are concurrent rather than longitudinal. The major study by De Hirsch et al (1966) made use of an abbreviated version of the Bender Gestalt and found a correlation of .56 between kindergarten Bender Gestalt and reading two years later. The author's own study found a somewhat lower correlation of .45 between Bender Gestalt and reading performance nine months later; stepwise regression, an arguable procedure, showed Bender Gestalt entering the regression first, well ahead of a simple measure of verbal I.Q. On balance, thus, there were firm grounds for using the Bender Gestalt in the present study, especially as few other measures of visual perception or integration have been so fully developed and tested.

The test used here is similar to that in the De Hirsch study, only six of the nine forms being employed; certain simplifications were made (as described in the Appendices A) for the sake of the four-year-old children who would be using the test in the nursery battery. The relevant Appendix explains the scoring procedure in the case of omitted diagrams or partial omissions.

8. Self-picture test. The originator of the draw-a-person test, Goodenough (1926), claimed that the test provides a reliable non-verbal measure of intelligence. In particular, drawings by young children were said to have an intellectual rather than aesthetic content and were determined by conceptual development. Her test was scored on a 51 point scale. The test was revised by Harris (1963) and the scoring altered to provide for a 73 point range for the draw-a-man and a 71 point range for the draw-a-woman picture. Harris also established broad age norms. The significant difference in the Harris approach was to see the skills tapped by the test in terms of intellectual or conceptual maturity rather than intelligence per se - specifically, as the ability to form concepts of an increasingly abstract character, to discriminate likenesses and differences, to classify and to generalise. He puts forward the hypothesis that in the early years of childhood the child moves increasingly towards a better representation of what he or she sees, and only later on do personality factors and creative abilities come to the fore in the drawings. Harris also found a relationship with numerical aptitude for kindergarten children.

Regrettably, the effect of nearly 30 years use and misuse of the original Goodenough test has been to create the impression that the test in its present form should still be regarded as a measure of intelligence. As Yule et al (1967) point out, the inter-rater reliabilities, though relatively high, are not such as



to ensure the accuracy of this test as a measure of I.Q. Moreover, the correlation of the test with the short W.I.S.C. is unacceptably low for the 9 to 11 year olds used in their sample.

Overall the reliability of the test is seen as satisfactory. Anastasi (1976) reviews various studies and finds high inter-scorer correlations and reasonable test-retest figures; of particular importance was the finding that re-administration of the test to groups of kindergarten children on consecutive days revealed no significant difference in performance. One factorial study found that reasoning, spatial aptitude and perceptual accuracy showed the highest correlations with the test.

The view in the present study was to see this test as in some degree a measure of mental maturity, in the cognitive or conceptual terms suggested by Harris (ibid). The Harris protocol was however modified somewhat to make it more suitable for the young sample and a combined protocol drawn up to take account of either male or female figures. Extra points are scorable for the presentation of intimate items such as the navel, nipples, etc. There are 54 scorable items in the present marking (see Appendices A), and to avoid the problems that seem inherent in asking young children to draw either a woman or a man each child was asked to draw a picture of itself. For the very small minority who could not grasp this concept, they were asked to draw "a picture of a man, or a woman" and their score was reduced by one point.

9. Piagetian test. This test was seen as offering a general indication of the child's developmental level, in Piagetian terms. It was not hypothesised that there would necessarily be a high relationship with reading performance at reception class level, though it was hoped that the test would show satisfactory correlations with other tests of cognitive development.

It would be impossible even to summarise, in a few pages, the vast range of Piagetian studies which might be thought related to the themes of development of nursery age children. Flavell (1963) is a standard text and there are many important volumes such as that by Sants and Butcher (1975) which present Piagetian thinking. It is less often that theoretical criticisms appear, such as those voiced in Tanner and Inhelder (1959), Toulmin (1971) and Ammon (1977), although there have been many experimental research caveats such as that demonstrated by Bryant (1972).

It is the research studies that are of particular relevance here. Modgil (1974) offers a broad review of this area, while research on acceleration studies - what Piaget termed 'the American question' - is summarised in Strauss (1972) and Sigel and Hooper (1968), among others. The ultimate focus, in terms of the present research, is on the extent to which young children follow an

ordered Piagetian developmental progression, and the relationship of levels of development to academic attainment. Flavell (1972) shows the methodological and conceptual complexities in the research interpretation of stage-sequence theory. One of the limited number of longitudinal studies in this area is that of Versey (1974) who followed 93 six-year-old children, in four cohorts, over an 18 month period and found support for a constant order of succession in the accomplishment of a wide range of tasks.

A larger number of longitudinal studies have looked at the relationship between Piagetian developmental levels and later academic attainment; Dimitrovsky and Almy (1975) followed 121 children for two and a half years and found that the kindergarten levels of conservation of number did distinguish sharply between good and poor readers (nearly all the conservers turned out to be good readers, but only half the non-conservers); they linked this finding to Elkind's theory that reading involves the ability to decenter perception, an ability which could enable children to recognise that one letter can represent more than one sound, and that the same sound can be represented by more than one letter. However, there are many methodological doubts about this kind of study of academic attainment. Frequently cognitive performance is not taken into account in a way that would show whether level of conservation, for example, adds significantly to the prediction from the cognitive scores alone; the role of other factors like visual integration is also usually omitted. In the field of mathematics Mpiangu and Gentile (1975) examined the Piagetian claim that conservation is a necessary condition for all rational activity and that arithmetical thought is no exception to the rule; they concluded that this was not shown to be the case, and suggested that conservation of number and arithmetical operations were conceptualisations that developed simultaneously. Studies such as these point to the difficulty in interpretation of the effects of Piagetian development on academic attainment.

For this research a composite test was devised with four sequential areas of assessment. The child was initially given three tasks requiring recognition of 'same' and 'different'; if it failed the last of these, or at any subsequent point, the test was halted. Then followed a series of number conservation tasks; on each occasion the child was asked (after E had extended or shortened the length of one row), "Have we still got the same number, or not?" However, no further attempt was made to disconfirm the child's belief, as is the normal practice, since it was felt that four and five-year-old children would be particularly susceptible to experimenter influence. The third area assessed the child's ability to perform seriation operations with a set of wooden blocks. The fourth area was a test of multiplicative classification, based on a study by Lovell et al (1962). The rationale for this ordering of tasks was based on Versey (ibid) who noted 87 per cent conservation of number at the start of testing his cohorts and an average 94 per cent 18 months later; 50 per cent seriation

success at the start and 95 per cent at the end; and an averaged 20 per cent multiplicative classification at the start, compared with 48 per cent at the end.

10-13. Sub-tests of the Wechsler Pre-School and Primary Scale of Intelligence. It is generally accepted that cognitive skills play a part in academic attainment, although the relationship is neither consistent nor simple. There are many contributory factors such as motivation, personality variables, the quality of teaching, teacher-pupil interaction, home pressures or apathy, and peer pressures, which likewise influence a child's attainment in any particular area of school work. Even the response to an intervention programme may be linked to the child's cognitive functioning, or to the parents' perceptions of those levels.

Cognitive assessment presents its own set of problems. The most practical of these is the choice of a test from among a wide range of highly professional instruments. The Wechsler Pre-School and Primary Scale of Intelligence was selected because it offered a range of interesting sub-tests with a floor just below the youngest age level of children in the study. However, as with any cognitive scale, there are particular problems in its employment. Maxwell (1972) found a considerable difference in the pattern of responses of good and poor readers, when given the WPPSI battery. Chance estimation of data in a statistical exercise revealed a marked and consistent difference in correlation between the sub-tests for the two groups, and this in turn led to the finding that for the good readers the ten sub-tests factored into only two main factors, compared with the three main factors needed to account for most of the variance for the poor readers. Hardy et al (1976), working with the Wechsler Intelligence Scale for Children, which contains many of the same items as the WPPSI battery, noted that in an immediate re-administration of the test to a sample of inner-city seven-year-olds, a number of the items were answered correctly on the second occasion. The children were asked about their answers and it was found that in many cases incorrect answers were given owing to misunderstanding of the questions rather than to lack of knowledge of the answer. It can of course be argued that while Hardy et al are right to point to the class bias inherent in the cognitive test situation, it is not very different from the average school situation where middle class values and expectations are often to the fore. Cognitive tests based on the same assumptions are likely to be better predictors of school performance than are tests which are geared to a more unbiased assessment of the inner city child.

Sattler (1974) offers a thorough methodological review of the WPPSI test, its advantages and limitations. He cites many of the studies which have examined the reliability and validity of the test and its administration; one fault is the lack of an adequate floor in some of the sub-tests. Of particular

interest are his own comments and quotations from other workers on the merits of individual sub-tests.

In the present research it was clearly necessary to obtain some fair indication of each child's cognitive skills. But it was also felt unnecessary to carry out a complete test of these skills. This was for two reasons. To conduct a full test on each of two occasions would have demanded a large expenditure of time with little indication that this would be of much additional value for the goals in this research. Secondly, any single battery of these sub-tests has its own idiosyncracies; such tests are not uniformly related to the cognitive skills which are thought to underly the criteria of reading and mathematical attainment.

It was thus decided to rely on four sub-tests of the WPPSI battery.

Research into the merits of this practice is offered by Silverstein (1970a, 1970b) who reports a series of studies in which data from the test manuals of the WPPSI, WISC and Wechsler Adult Intelligence Scale were utilised to derive evidence on the correlations between the 'short forms' of these tests (from two to five sub-tests) and the full-scale tests. His figures showed that such full-scale correlations ranged from .82 to .83 for different combinations of two WPPSI sub-tests, and from .903 to .906 for various combinations of four sub-tests. Thus it was felt that the use of four sub-tests in the present research would not present an unduly limited assessment of the children's cognitive skills. The choice of these sub-tests demanded much consideration.

Sattler's description and evaluation of each sub-test were of help here. The four tests chosen for the research were information, sentences, picture completion and block design.

His report on the information sub-test says that it appears to provide an adequate sample of the information acquired by a person who has had the usual opportunities in society, reflecting the richness of early environment and cultural predilections. For the purposes of the present research this test was seen as a basic measure of the child's verbal/cognitive past experience. In his discussion of the sentences test Sattler reports that the test's reliance on immediate recall and attention means that it is seen as a memory test, but one which is also related to verbal facility. Although this is only a reserve or supplementary test in the battery, and although it is not a fine-grained discriminator (for example, a child who might hesitate and then refuse to attempt a sentence because he or she has forgotten one key word will lose all the points for that sentence), it was chosen for the present research because it is thought that short-term memory plays a considerable part in early reading, when learning to recognise words or letters.

Sattler's summary of views on the picture completion test indicated that

it is viewed as a test of ability to differentiate essential from non-essential details, requiring visual organisation and visual memory. Again, such a description seems well related to the reading goal in the current project. The fourth sub-test, block design, is seen by Sattler and others as requiring the perception and analysis of intricate designs and the transposition of these insights to other blocks so as to reconstruct the pattern; qualities identified by this test include visuo-motor coordination, perceptual organisation, spatial visualisation, and concept formation involving both analysis and synthesis. Cronbach (1961) has described the Kohs Block Design test, from which this test is derived, as a good non-verbal measure of analytic and synthetic reasoning with a wide range of difficulty. Although taking considerably longer in administration than any of the other sub-tests, it was seen as an essential constituent of the battery of tests in the present research, mainly because of its evident ability to assess deeper cognitive skills.

A few further points need to be made about these tests. Since it was not planned to compare the children's scores with the test norms, nor to make detailed use of any global test figures, only the raw scores were used. These were employed in the analyses as independent variables together with age, the latter being measured in half-month units (the WPPSI manual uses three-month intervals).

The reliance on raw scores also enabled other minor breaks with the strict test protocol. In the information sub-test if a child volunteered some 'name' when asked to give the name of an animal the question would be rephrased: "What kind of animal do you like?" The manual suggests asking such a child: "What is (Rusty)?", which assumes that the child would volunteer the name of a domestic pet. In the present research almost any 'name' was volunteered by children who did not understand the initial question. For the sentences sub-test if a child volunteered an ethnic 'have' in place of 'has' it was scored as correct (in 'Mary has a red coat'); however 'had' was disallowed, as this was clearly a change in meaning.

In the picture completion sub-test an alternative presentation of the question "What's missing from this picture?" was the colloquial "What's been lost (losing) from this picture?" (initially volunteered by a child of West Indian descent). A more substantial change was the decision to accept 'back wheels' as one of the valid answers to item 15. The rationale for this and the statistical analysis carried out to justify that interpretation are described in section 6.50. Permission was obtained from the Psychological Corporation of New York to reproduce the item for the purpose of that discussion.

In the block design test the fourth item requires the presentation of two blocks in a diagonal design. For the four-year-old children it was often fortuitous whether their first attempt pointed the blocks in the correct or incorrect diagonal direction; research by Olson (1970) showed that for very young children

the diagonal is still a somewhat shadowy concept. It was thus decided to allow either presentation in the nursery battery, but to require the correct presentation for the reception class children.

14. Distractibility. The assessment of a child's level of distractibility is difficult and uncertain since it is context specific and subject to all the problems of other forms of personality assessment. Nevertheless it is seen as important. The subjective assessment of distractibility in the author's previous research yielded a correlation of .43 with the combined reading score nine months later.

For the present research it was decided to use both a subjective assessment, made in the first and second halves of the test session, and an instrumental assessment based on apparatus tailor-made by the Technical Department of London University's Institute of Education. Electronic timers and other materials were borrowed from existing equipment. The instrument was programmed to make three different kinds of noise - a bell, a chime and a buzzer, at regular intervals which could be pre-set; in this case three-second intervals were used. The instrument was housed in a plain wooden box and placed on the floor close to the testing table. The starting switch and sound alteration controls were held unobtrusively under the testing table.

#### Needs assessments

Although it was hoped that the parent interviews would present a fairly reliable picture of the home behaviours which contributed specifically to later reading and mathematical performance, it was also felt necessary to carry out an assessment of the child's state of needs, as defined by Maslow (1945). The structuring of the interviews was such that no emotionally 'difficult' questions were put to the parents. But in view of the widespread evidence on the links between unstable home situations and poor academic attainment (Davie et al, 1972, among others), it was felt that such factors should also be taken into account in assessing the home background to the children's academic attainment.

In the author's previous research the teachers were asked to fill in forms which yielded data on four of Maslow's needs - physiological, safety and security, love and belonging, and self-esteem and worth; only two of these variables yielded significant correlations with reading performance nine months later - the need for safety and security (.33) and the need for self-esteem and worth (.29). Details of the form which the teachers in the present research were asked to complete appear in the Appendices A. In virtually all cases it was the teachers responsible for the children at the time of assessment who were asked to fill in the forms; only in the relatively small number of cases where

children had left their schools for other schools shortly before the mid-test battery, were the original teachers asked to fill in the forms.

These assessments were carried out at the same time as the nursery and mid-test batteries were administered; in effect, thus the initial needs assessment can be linked to the parent interviews carried out at the same time. The second assessment can be linked to the parent programme data; although not co-terminous with the programmes, this second assessment can be hypothesised to represent the child's state of needs in the period prior to the teachers' filling in of the forms, and is thus to some degree intermediate between the nursery and mid-test batteries.

#### 4.312 Mid-tests (reception class)

The 10 cognitive tests (numbers 5 to 14 in the above list) formed the mid-test battery and were administered to all the children approximately three to four weeks after entry into reception class. (The few exceptions to this schedule are discussed in section 6.10.) As described above, the needs characteristics were also re-assessed at this stage.

#### 4.32 Parent interviews ...../

#### 4.32 Parent interviews

The interview with parents of the sample children was clearly of major importance. Any judgements on the apparent effectiveness or otherwise of a parent intervention programme would need to take statistical account of the parents' initial behaviours. Thus a parent who was already doing a great deal of pre-school work with her child at home would be more likely to participate and remain in a parent programme than would a parent who saw little purpose in the exercise; success in the one case and failure in the other could well be expected. Only a sensitive design and analysis, based in part on the parents' initial levels of home behaviours, could hope to unravel the relative success or failure of the programme in each case.

In essence it was hoped to identify and measure those aspects of the home environment which are thought to contribute to early academic performance in school entrants. It was decided that emphasis would be laid on the determination of behaviours, with attitudes assessed only as a secondary measure. A great deal of evidence (Cannell and Kahn 1968, Cox and Rutter 1976, and Blum and Naylor 1968) points to the serious problems of low reliability and validity for interview measures, these problems being accentuated in the case of attitude measures as opposed to factual information. For example, evidence from Hewison (1980), gathered in an area related to the present research, showed that interview information on parent behaviours bore a good relationship to their children's reading performance whereas parental attitudes and other home literacy factors often showed little relationship to that performance. The Thomas, Chess and Birch project (1968) found that an apparent homogeneity of attitudes towards child-rearing was not born out by a homogeneity of practices.

Alternative methods of assessing parent home behaviours were given brief consideration, but the evidence of major reviews such as that of Cannell and Kahn (ibid) and a Newson and Newson discussion (1976a) suggest that few other methods can compare in effectiveness with the interview for obtaining the kind of personal data sought in many studies. The only alternative which is a real contender within the area of parent home behaviours is the combined questionnaire/observational study which can be undertaken using instruments such as the Caldwell Home Inventory (Elardo, Bradley and Caldwell, 1975; one of the few English trials appears in Madrid-Jimenez, 1977); this is an excellent instrument and could possibly have been adapted for the present study, but its administration takes an estimated hour on average, whereas it was hoped to spend an average of half an hour per interview in order to complete the planned 200 interviews on schedule.

Among the most noted of English studies on parent behaviours and attitudes



are the series by Newson and Newson (1963, 1968, 1976b). This couple consider that it is necessary to assess both behaviours and attitudes, as these complement each other and thus offer greater reliability. The Newsons' development of the interview protocols for their studies and their sensitive handling of respondents present many useful ideas, although they have limited recourse to hard statistical analysis. Another noted study relying partly on parent interviews was that of Thomas, Chess and Birch (ibid), who studied in depth the development and home environment of 141 New York children. Although Thomas et al used regular interviews as part of their longitudinal study, Cox and Rutter (ibid) point out that a great deal of information about family life and relationships can be obtained even from a single interview with one parent. Apart from the Thomas et al study there have been several other American developmental studies which have followed samples of children, and their parents, from birth onwards; in Britain the Douglas (1964) and Davie et al (1972) reports assessed certain home factors in large scale longitudinal studies. A recent study by Kellaghan (1977) looked at home environments in more detail, particularly in relation to specific academic performance of the children.

In the wider field of research on the interview process, little attention has been given to the methods and quality of parent interviewing as such. Bell (1969) offers a small-scale study from within the Birmingham E.P.A. project, but extensive methodological discussions such as those of Smith J(1972) and Richardson et al (1965), while offering many ideas on general interviewing concepts and the variety of available techniques, have nothing to say on the rapidly growing area of interview research into parent home environments. It is thus a relatively uncharted field, methodologically speaking, in which the interviewing work for this study was carried out, although the various authorities on interviewing do present a wealth of information on the basic processes and also stress the practical do's and don'ts for the interviewer.

The protocol used here was developed on the basis of interviews with twelve volunteer parents from a housing estate surrounding a modern nursery school within the local authority area. In retrospect it can be seen that a wider pilot sample, perhaps 30, may have indicated the need to expand some of the items which later proved to have too limited a range or power of discrimination, when used over the full sample.

The characteristics of the interviewer are always seen as of considerable importance in the intimacy of the personal interview. In the present case these characteristics were the following: E is of European ethnic origin, middle-aged and male. Only in the case of Asian households was the male characteristic a handicap; in such cases the interviewer would usually only be admitted if the husband was at home, and it was normally the husband or other male figure who took the lead in answering the questions or discussing home behaviours.

The interview schedule and a fairly typical wording of that schedule are presented in Appendices B2 and B1. It can be described as a standardised interview, administered in semi-structured form; in other words, the same questions were put to all respondents, but the structuring of the questions was to some degree adapted to suit the circumstances and cultural milieu of the respondents.

It was decided to start interviewing parents at the same time as their children were being assessed in the nursery classes. Thus when E started work in a class, letters would be sent out by the school to all parents. A typical letter is reproduced in Appendix B4. The letter enabled a parent to refuse permission, but otherwise the assumption was that no reply indicated consent. This concession was made by the administering Authority (as opposed to the far more restrictive method of requiring positive consent, which usually produces a heavily biased sample of respondents), since E had given the assurance that no intimate personal questions would be put to the parents.

Out of more than 210 parents receiving these letters at the six schools there were only six straight refusals, with a few other refusals for understandable reasons such as the imminent birth of a baby or a pending marriage break-up. Two husbands refused their wives permission to take part in programmes; in the one case the mother approached E later, during the mid-test period, and asked whether a full assessment could be carried out on her child as she was gravely concerned about his apparent backwardness.

#### 4.321 Areas of inquiry

The interview schedules in the Appendix and especially the interview coding (Appendix B3) indicate the ten broad areas covered by the instrument. The first of these, the ethnic group(s) of the parents, was coded on the basis of information provided by the parents during the interview; no questions were put on this issue.

The parent(s) reading behaviours were studied in some detail, altogether six questions being aggregated to provide this score; the main question here concerned the frequency with which either parent read to the child, the question being approached indirectly so as not to elicit an answer reflecting the ideal rather than the reality; other questions concerned membership of a library and similar 'behaviours'. The child's language environment was also studied in terms of behavioural criteria, such as whether the parent made up stories to tell the child, and the frequency with which the child watched Playschool. The parent's reading attitudes were probed in an attempt to determine whether the parent saw a role for herself in starting the child's reading process, or whether this was seen as entirely the task of the school. Two questions on the issue, at different

points in the interview, were used as a cross-check to assess the strength of this attitude.

The mathematical environment was more difficult to assess, as parents are less aware of these behaviours (the fostering of number and mathematical concepts) than they are of reading behaviours. Four items, one of them assessed from the general discussion rather than from a specific question, were used here. The parent's mathematical or 'number' attitudes were assessed in a question parallel to the two questions on reading attitudes.

In an endeavour to quantify the 'motivational' situation at home, questions were put on two issues. The first set examined the parent's view of the degree to which the child could be persuaded to behave nicely to other children (sharing, etc.), and to co-operate in performing household tasks or helping in other ways. The second set of questions probed the parent's views on nursery schooling and its purpose, and on how the parent saw her role and that of the school in encouraging the child to 'learn how to work' in the school setting - issued concerning, for example, the child's willingness to tackle school tasks or to persevere with them.

The last two areas assessed the approximate time which the child spent viewing television, on week-days and week-ends, and the parent's behaviour and attitudes in controlling the child's viewing in any way.

The limitations of the instrument meant that four areas could not be covered by formal questions, although these have been described as relevant to the child's 'educational' home environment. The first was whether the child was minded for any part of the day, or whether he or she had been minded for any long period in the past; some of the parents who broached the subject and volunteered information on this were clearly sensitive to the widely publicised criticism of the poor quality of many child minders. Whether there was a man present in the household (when the father or other male was not present at the interview) was also something that could not be clearly ascertained without a specific question; this could have proved embarrassing for single mothers, particularly those supported by social security on the grounds that they are unsupported by any men friends. Nor was social class ascertained. This was for several reasons. Although Newson and Newson (ibid) have studied in great depth the effect of social class on parent behaviours and attitudes, it is clear from the examples cited in their reports, and elsewhere, that what distinguishes one class from another (apart from the Registrar-General's occupational divisions) is a rather loose set of sub-cultural behaviours rather than any indefinable 'class' creed which somehow imposes a particular behaviour pattern. Even the sub-cultural behaviours identified by the Newsons do not appear uniformly across a class and there is much overlap. The important study by Douglas (ibid) emphasised the serious limitations of the Registrar-General's classifications in terms of educational attain-

ment. It was thus seen as more relevant to look at the behaviours themselves rather than to complicate the analysis by a gross categorisation in terms of occupation. In other words, the importance of social class is not denied but it was felt that other specific behavioural instruments were more relevant in the present research.

The arguments for the omission of these three areas from the interview protocol are not clearcut and there are conceptual and ethical considerations on both sides. The fourth area is less arguable; the level of home 'stimulation', such as parental encouragement of specific pre-academic skills and motivation to succeed at tasks, has been shown by many workers (Douglas (ibid), Deutsch 1964, Vernon 1969 and White 1973) to be of significant importance for the child. Instruments such as the Caldwell Home Inventory and the Wolf (1966) research protocol have been shown to be effective in assessing this level of home stimulation. As was pointed out earlier in the section, limitations of time prevented the addition of such instruments to the present interview protocol, which looked more specifically at pre-reading and pre-mathematics behaviours.

#### 4.322 Controls and scoring

There were several forms of statistical control, apart from the coding of the interview items. Six of the interviews were tape-recorded and these were listened to and marked on an interview protocol by a colleague of E (details appear in section 6.50). Permission to tape the interviews required careful negotiation, including the assistance of the nursery teachers in the choice of parents with sufficient confidence to agree to such a request. The nature of this preliminary selection of parents meant that the recorded interviews were by no means typical of the sample, let alone random. The same problem of lack of randomness arose with the method used to check on E's assessment of the child's television watching hours. Seven parents were selected, again an untypical sample, on the basis of the frankness of their answers during the interviews; at the end of these interviews, after E had made his assessment of the watching times, the parents in the seven cases were informed that E had made a rough assessment of the hours, on the basis of the interview discussion about the kinds of programmes seen by the children on week-days and week-ends. The parents were not initially given E's figures but were asked to calculate the daily viewing for both week-days and week-ends. Details of this comparison are also provided in section 6.50.

The decision to aggregate the scores from several items in order to arrive at an overall score for a particular area was taken in the light of competing viewpoints. On the one hand Galtung (1967) argues that additive indices can

lead to various statistical problems - for example, adding an item with a wide range of scores to one with a low range can impose numerical values on variables whose structure is uncertain; Gardner (1975) considers that this problem is particularly serious in attempting to combine attitudes. Galtung suggests that at most the values of additive items should be restricted to two (is, is not) or three (yes, don't know, no). On the other hand one can argue, firstly, that a large number of separate items would cause impossible complexity in the statistical analysis, and that some combination is essential to gain insight into the research issues; and secondly, that there may be good conceptual reasons why a particular item of importance should be rated more highly than an ancillary item. For example, whether one or both parents read to the child is a useful behavioural indicator, but it is less important than the quantification of how often the child is read to by its parent(s) each week. Two points were allocated to the first item, and four to the second, in combination with other behavioural indices in the reading area.

A methodological problem inevitably arises when parents are interviewed about home behaviours which many parents would recognise as being related to the child's later academic performance. To what extent might E's questioning on these issues have prompted an increase in the behaviours discussed? There appears to be no way in which this can be determined. On the positive side it might be argued that since all parents had a standardised interview, the effects (within the normal range of interaction across parents) would be evenly spread across both the experimental and control parents.

The final control problem concerns the reliability and validity of the instrument, in wider terms than the limited exercises described in the last few paragraphs.

These issues are discussed in detail in sections 5.14, 5.15 and 6.50.

#### 4.33 Parent programmes

The goals of the parent programmes have been discussed at some length in earlier sections, especially in section 4.10. Briefly, it was the aim to provide parents in disadvantaged areas with a fairly structured programme in early reading or early mathematics, to enable them to become aware of and gain confidence in their ability to influence the educational development of their children. The programmes laid great emphasis on the need to relate any attempted activities to the developmental level and continuing interest of the child.

Prior to the start of the main programme, permission was obtained from a nursery school within the same urban context to run pilot reading and mathematics programmes for the 12 parents who had been interviewed in the pilot interviews. The parents were randomly divided into reading and mathematics groups and each programme ran for eight meetings, at intervals of two or three weeks. There was some loss of participants, for reasons such as taking up employment, the child moving to another school, or decline of interest in the programme. Half the parents were still present at the final or pre-final meetings.

The organisation of these programmes and the parents' comments on the tasks and activities provided valuable guidance for the preparation of the main programmes, which started about a month after the conclusion of the pilot series.

For a variety of reasons it was decided to keep to the fortnightly schedule for the main programmes (except when interrupted by school holiday periods). This was seen as a compromise between weekly meetings, which it was felt would be too demanding for most parents, and monthly meetings which might be too easily forgotten. However a small number of groups were allocated to a six-weekly meeting schedule, with letters sent to all parents in advance of each meeting, as a means of assessing the effectiveness of holding meetings at such long intervals and the consensus of parental views on this schedule. The fortnightly meetings lasted one hour and the six-weekly meetings two hours, the latter groups having a break for tea. The meeting agendas for the six-weekly groups were the same as those given to the fortnightly groups, but clearly a number of matters could not be dealt with in the same detail, since overall the six-weekly groups spent only two-thirds of the time at programme meetings as did the two-weekly groups.

The four larger schools each had between four and six programme groups, and the two smaller schools three groups each. There were up to six parents in a group and no group started with less than four parents on the roll. During the formation of parent groups a somewhat greater number of parents were

allocated to reading than to mathematics groups. In total, 25 groups were formed, 14 of these for the reading programme and 11 for the mathematics programme, with reading groups generally larger than the mathematics counterparts.

Two meetings were organised for the morning and two for the afternoon nursery sessions (except in the case of six-weekly meetings, which covered an entire nursery session). The schedules were arranged so that E spent a full day every two weeks at each of the four larger schools, and two half days at each of the two smaller schools. Six-weekly meetings were organised during the intervening weeks.

Each school willingly provided a meeting venue. Initially the possibility had been considered of holding meetings in nearby church halls or other community centres, but since this would involve extra travelling for the parents (usually on foot) and would also demand considerable effort to organise heating and other facilities, the schools were approached for their assistance. In four cases the meetings were held in the entrance foyer of the nursery class and in two other cases in the school's medical room. E prepared the meeting venues, arranged the display of posters on the nursery doors, and also posted a list of the names of parents due to attend that session's meetings. As far as possible all meetings were held at regular intervals and every parent was issued with a programme calendar in advance of the first meetings.

When it became apparent, in the course of the programmes, that many participants planned to take up employment when their nursery children entered reception class, it was decided to end both programmes at the 8th meeting, just before the school holidays, since a considerable number of sample children would be going into reception class in September (1977) and the numbers in most programme groups would be likely to drop drastically. Accordingly some of the important features of the agendas planned for the final six meetings (9th to 14th) were abstracted and the content of the sixth, seventh and eighth meetings altered accordingly. It was explained to parents, and emphasised repeatedly, that both the extra material and all the original material were not intended as pace-setters but rather as information for parents to apply when their children were ready and interested in tackling the particular activities. It was stressed that these were parent education programmes rather than recipes for any particular child.

Appendix C2 contains the basic forms sent out to parents. The wording of these forms indicates the specific approach taken towards the participants at the outset and the cautions voiced about trying to accelerate the children's progress unduly. The introductory letters and materials for the reading and mathematics programmes emphasised the initial goals of these programmes, such as (for reading parents) reading regularly to the child and joining the local

library, or (for mathematics parents) giving the child a set of 25 Unifix play blocks (interlocking) and suggesting that these be linked with finger matching to help the child develop a sense of number within a spatial context.

All the materials referred to in the meeting agendas and other programme sheets, including the Unifix blocks, were provided by E. Although the preparation of materials is often undertaken by participants in intervention projects, under the guidance of programme organisers, it was felt that explaining and practising such tasks would take up an undue amount of meeting time; the quality of what was prepared might vary so much that some of it would be of little use; and, given the uncertainty with which many parents took part in the programmes - as a matter of trying them out rather than of conviction - it was thought that some participants might lose interest if there were additional demands upon their time at home, apart from what they were asked to undertake with their children. This is admittedly a debatable issue, for there are good arguments in favour of the creative aspect of such preparation and its stimulating effects on the people concerned.

The design of the reading programme was based not only on a wide study of literature in this field, ranging from imaginative though basic books on how to help infants to read (McMillan, 1973) to the more sophisticated works such as that of Mackay et al (1970), whose Breakthrough method of teaching children to read is one of the major reading schemes followed in British infant schools today. Other reading schemes were also examined. Reception teachers at all the schools were consulted at some length about the methods they used in teaching reading to children. Advice was also requested from other educational sources.

It was realised that no one programme could meet the needs of all participants, nor indeed would any one programme necessarily offer the kind of individually-gearred sequence of activities most suited to a particular child; however it was hoped that each parent's knowledge of her own child would enable her to take from the programme those elements most suited to the child's needs or capacity. There were other problems too, such as the limited level of literacy possessed by some parents; it was hoped that the programme might even help to strengthen the reading foundations of such parents, to some small degree.

A sensitive issue was the question of whether the reception teachers would all approve of the idea of encouraging parents to start their children on the road to reading, even at such a basic level. No teacher objected, although a few had doubts as to whether any such programme could prove effective.

The part played by parental behaviours in the educational and linguistic development of the child was discussed at the start of the first and subsequent



meetings of every group (both reading and mathematics), and the need to understand the child's point of view and developmental level was stressed. These discussions of general ideas were seen as particularly important. A summary of a number of the key points is reproduced in table 1 overleaf. Many other developmental issues were discussed when raised by parents or when they were relevant to a particular activity.

At the initial meeting of each reading group E discussed some of the ways in which parents influence a child's verbal and reading development and pointed to the overwhelming importance of verbal and reading fluency in later life. The difficulty faced by a child in coping with the seemingly incomprehensible hieroglyphics of words and letters was discussed, together with the importance of the environment as an easy avenue to early word recognition. Problems such as 'barking at print' and other incorrect reading behaviours were reviewed. At this and a number of the later meetings the general principles underlying the reading process were described in an easily digestible form; these principles were drawn inter alia from the ideas of Frank Smith (1971, 1973, 1975) and of the various authors in the selection edited by Jessie Reid (1972).

The contents of the reading programme are given in the agendas for the eight reading meetings, reproduced in Appendix C3; Appendix C4 contains some of the ancillary programme material. Two typical examples of the ancillary material appear in the next few pages. The first illustration reproduces in full size one of the composite alphabet word cards from the set given to parents; other cards in this set included the individual alphabet words (for matching and later recognition), a selection of high frequency words, and single letter cards. Cards of substantial thickness were used for the set, which was designed by E and printed for the reading programme parents with the help of financial support from the University of London's Central Research Fund and the London Institute of Education's Minor Research Support Fund. The second illustration reproduces two pages from the second reader given to programme parents. The words and drawings in the first reader (Appendix C4) were more elemental and much larger.

The summary below of the points covered in the first five programme meetings gives an indication of the kind of activities suggested to parents, with the aid of materials where necessary. (Since the agendas for the 6th, 7th and 8th meetings contain additional material which had originally been planned for a later stage in the programme, they are not typical of the sequence or pace of the first five meetings.)

Meeting 1: Introductory discussion; issue of alphabet word and letter cards, with matching picture cards; discussion of the use of local libraries; activities with the initial reader (issued in advance); identifying a few key words

Table 1.      Summary of key points emphasised during programme meetings

- a. Parents' role in the development of the child:
  - physical health - sufficient sleep/nutrition/attendance at child clinics
  - language - the whole talking environment; use of 'adult' language
  - knowledge of the world - outings
  - understanding of what is happening - answering child's questions
  - child's thinking - the kind of questions parents ask the child
  - moral awareness - fairness to others
- b. Society's attitude towards parents:
  - financial support - child benefits, family supplements
  - taxation - discriminatory policies over many years
  - maintenance during schooling and non-university F.E. - grants minimal
- c. Parents as educators:
  - schools build on what parents have already fostered in the child, such as language, experiences, skills, attitudes
  - importance of good relationships with teachers and cooperation in joint task
  - parents' role in the school years - individual help for their children
- d. Differential development:
  - children (programme and otherwise) at different ages, stages of development
  - different rates of learning to talk, walk; height spurts; late developers
  - danger of 'competition' with other parents; each child at its own pace
- e. Praise/criticism when working with child:
  - positive encouragement through praise; damaging effect of criticism, loss of interest; help child to discover mistakes in 'joking' way; try again
- f. Particular techniques:
  - sitting at low level; 'games' rather than work; omit activities if child has been naughty; end before the child tired, and promise more later
  - reward with praise and a cuddle for effort rather than mere success
  - danger of boredom if activities too repetitious
  - if child not ready for one activity, try another; pick and choose; keep other activities for later and hide materials
  - use of the typed agendas; item difficulty levels of (i), (ii) and (iii)
- g. Parents' contribution to the programmes:
  - E looking for parents' ideas; value of exchanging ideas within the group
  - research will aid in determining how to present ideas to other parents
- h. General:
  - meeting schedules; keeping to agendas, within limits
  - special arrangements when meetings missed unavoidably
  - value of continuing to attend even if child not up to a particular point









in the reader and in other books; finding the same words on different pages; matching a few cardboard words with the labelled pictures on the alphabet card; general discussion of the principles of reading and learning to read.

Meeting 2: Distinguishing words from letters; knowing own name letters; letter matching with first letters on word cards; fostering the child's interest in 'family' words and names; directionality in books; exchange of programme reading books; the holding of a pen and the two basic directional rules in writing; copying or tracing ones own name; nursery rhymes.

Meeting 3: The alphabet song; choice of reading matter at the library - examples of good modern books; word lotto; flash words from the initial reader; matching words; 'writing' difficult letters in the air; preparation of child's own story book, with a booklet of extra large pages and fibre pens issued to parents; pictures drawn by child and sentences written by parents in child's own words; copying of these sentences.

Meeting 4: Nursery rhymes and rhythmic pointing; recognition of words in parents' environment; parent writing five or six key (environment or family) words and child learning to sort and recognise these; recognition of words in isolation; learning of some letter shapes; I spy games based on sounds; extension of child's own story book; issue of second reader.

Meeting 5: Lotto games, using words and letters; more environment words (selected by parent, not by E); important clue in first letters; matching letters with words in the reader; finding the right letters; letter sounds; the concept of a 'sentence'; minimal importance of letter and word reversals at this age; spaces between words; making up verbal sentences.

At the start of each meeting the most important ideas from the previous meeting were reviewed briefly. Problems or successes in the carrying out of those ideas with sample children were raised and discussed by some of the parents.

For the mathematics programme much reliance was had on the study by Lovell (1971a) on the teaching of mathematics to children from kindergarten upwards, on the Soviet study by Taruntaeva (1971) on starting mathematics with four-year-olds, and on the guide lines for primary school mathematics issued every few years by the Education Inspectorate of the Inner London Education Authority. Reception teachers in the sample schools were also consulted as to the methods they used.

It was a particular aim of this programme to provide a sound foundation for mathematical understanding as well as some fluency in early numeracy. Accordingly the meeting agendas covered a wide range of mathematically-related topics, with sets being introduced at an early age. Appendix C5 gives details

of the mathematics meeting agendas, while Appendix C6 contains most of the ancillary sheets issued with these agendas. Two of these ancillary items are reproduced in illustrations on preceding pages in this section. One illustration shows some of the cards from a fairly extensive number set given to parents; apart from the basic numbers and matching symbols there were cards suitable for a modified domino game (numbers ran up to 20) and simplified 'fraction' cards for the last stage of the programme. This set, as with the reading set, was designed by E and printed with the help of the two university funds which supported the printing of the reading cards. The second illustration reproduces one of a number of activity sheets dealing with sets and simple number recognition. Other materials supplied to parents such as cardboard clocks, sets of cardboard coins, large dice and similar wares were typical of what can be purchased from educational suppliers, while materials such as logiblocks, balancing 'hangers' and other equipment were produced by E, with assistance from his wife.

At the introductory meeting the importance of mathematics in modern society was stressed. The ideas underlying the concepts of new mathematics were discussed, as well as the importance of parent-child number activities in the home. It was explained that mathematics is not a game with rules dictated by someone else, but rather a game in which the child makes the rules on the basis of her or his experiences with mathematical kinds of activities. The value of breaking tasks down into simple elements, and other desirable goals such as helping a child to find the right answer rather than providing it didactically, were also analysed. At the start of each successive meeting the activities carried out during the past two weeks were reviewed and other matters were raised about parental effectiveness in general.

Here too a summary of the main points covered in the first five meetings will indicate the approach to this programme:

Meeting 1: Introductory discussion; issue of the set of number and shape cards; simple counting activities; matching cubes with number cards; sorting (e.g. cubes by colour, cutlery by shape); sorting number cards by symbols; comparisons and groupings; experience with colours.

Meeting 2: Number play (nursery rhymes and finger work); ordinal counting; recognition of a few numbers and copying; sets and one-to-one correspondence; logiblocks; sorting by one criterion; sorting by two or more criteria; concepts in various fields - shapes, comparisons, spatial.

Meeting 3: Lotto; basic adding and subtracting concepts, more than, less than; sets and more than, less than; finger play add and take away; sorting and partitioning cards; intersection of sets and mapping; modified domino game; ordering by size; matching cube patterns; coins and matching.



Meeting 4: Counting games on various number sheets; comparative concepts; numbers and more than, less than; further ordering and matching; simple addition and subtraction with cubes; use of line cards and dice for number play.

Meeting 5: Counting in a circle; missing numerals; conserving quantity; joining and partitioning sets; sorting into sets and subsets; sequences; names of coins and recognition activities; following a drawn picture route.

There were two major problems of which E was particularly aware in organising these two programmes. Both the parents and the children, as dyads and individually, were at widely differing levels of development across the sample and there could be no optimum rate of progress in the programme. This created the difficulty that the running of a programme at a certain rate might be too fast and therefore threatening for some parents and children, or too slow and thus uninteresting for other parents and children. A solution was sought in providing programmes which covered a considerable amount of ground, while emphasising repeatedly to parents that the agendas and materials were also intended to be stored against the time when a particular child might be ready for certain activities, or alternatively that from each agenda only those activities should be selected which the parent thought would be appropriate for the child at its present level of development.

To make this selection of activities easier, most items on both the reading and mathematics agendas had sub-items in ascending order of difficulty.

#### 4.34 Child post-tests

The criterion variables on which the judgement of the effectiveness or otherwise of the intervention programme is based were decided upon at the start of the research. A battery of five tests was assembled.

1. Infant Reading Test. This test has been fully described in section 4.311. The identical test was administered near the end of the 1977-1978 academic year. The interval between the initial and final administration of this test ranged between 12 and 18 months and it is unlikely that there was any residual memory effect from the initial administration.
2. Southgate Reading Test. Although prescribed as a group test, this test (Test 1) was administered individually, as detailed in Appendix A5. Fumfrey (1970) writes that the test was specially devised for children in the early stages of reading acquisition; its manual gives evidence of satisfactory reliabilities and validities. The tests have been used extensively in national and local surveys. It can thus be seen not only as a key measure of reading attainment but also as a means of assessing the validity of the Infant Reading Test. The Southgate test was scored according to protocol.
3. Daniels and Diack Standard Reading Test (Sentences). This test was chosen as a companion to the Southgate test because it reflects another aspect of the reading process - the ability to read simple sentences with understanding. Fumfrey (ibid) points out that there are serious technical limitations in the manual for this test, including the absence of test reliabilities and validities; it is nevertheless a popular test and he points out that many teachers have found it to be most useful as a diagnostic instrument. (The Sentences test is only one of twelve sub-tests.) Two considerations led to a modification of the scoring procedure. Firstly, in terms of the research design some children had to be given the post-tests at ages below the minimum Daniels and Diack test norm of 5.2 years. Secondly, the author's previous research (Barker, 1976) noted that a fair proportion of less capable readers scored zero in the reception year if marked according to the Daniels and Diack 'all or nothing' protocol. The scoring was thus modified here to enable marks to be gained for partial success in reading a sentence. A further modification of the scoring system occurred during the testing itself. It was found that all children who came within reach of the final (20th) sentence scored full marks on this sentence. E then gave the sentence (but did not score it) to a number of children who had failed three successive items at an earlier stage and found that many of them also succeeded on this item. The item appears to be misplaced in terms of difficulty level. Consequently, although the 20th sentence was administered as a final morale

booster, the item score was not included in the scoring of this test. The test is described in Appendix A5.

4. Mathematics test. This test is the same as that described in section 4.311.

5. Piagetian test. This test is also the same as the test described in section 4.311. The Piagetian test was the only test to be administered on all three occasions - in the nursery, mid-test and post-test batteries. The conceptual issues underlying the scoring system have been discussed in that section. The experience in administering the test to the whole sample in the nursery and mid-test batteries suggested that there might be many children who, although failing on conservation of number, could succeed in the seriation test, or that seriation failure would not necessarily imply failure in the early levels of multiple classification. Accordingly the post-test Piagetian measure was scored (and administered) in such a way that two scores could be derived from the measure - one score on the basis of the original protocol, and a second score on the basis that failure at one level did not preclude continuing the test at the next level. The implications of this decision, particularly in relation to the final analyses, are discussed in greater detail at several points in chapter 6.00. The protocol for this alternative form of scoring appears in Appendix A3, immediately following the description of the original method of scoring.

## 5.00      A n a l y s i s

Tukey (1962) in a seminal study on the future of data analysis urged the importance of tackling statistical problems in a more realistic framework, if necessary evading deeper lying constraints, to explore novel approaches. Data analysis should seek scope and usefulness rather than certainty, he wrote; it should be willing to err moderately so that inadequate evidence might suggest the right answer more often; and it should use mathematical argument and mathematical results as the bases for judgement rather than as bases for proof or as stamps of validity. There were dangers in optimisation in terms of a precise and often inadequate criterion; more time should be spent on real research rather than on trying to convert a useful solution into the 'best' solution. Elsewhere Tukey (1977) emphasises the need for data exploration rather than restricting oneself to the planned analysis of an experiment.

Tukey's guidelines underlie much of the approach to data analysis in the present study. New approaches suggested by leading statisticians such as Wold, by lesser known innovative statisticians such as Hoerl and Vinod, and by the respected methodologist Cronbach, have been examined and further developed in the expectation that they will yield more interesting insights into the data than would conventional methods. The new approaches undoubtedly suffer from rough edges which only time and an application to other studies will reveal as needing modification. The extra year's research spent on developing a non-stochastic ridge regression technique has not only provided a by now well tested method for analysis of this study's data but also suggested that non-stochastic V-ridge is a major analytical tool alongside and often superior to ordinary least squares regression.

Applebee (1971) offers a critical examination of the fallibility of simple hypotheses and techniques when undertaking research into a problem such as reading retardation. There is no straightforward definition that can be offered of the problem itself. Attempts to identify a single cause or a set of separate causes all reveal serious shortcomings. The methodological orientation towards simple models does not fit the problem with which one is dealing; it is necessary to employ models which correspond more closely to the heterogeneity of the disorder, with more sophisticated methods of analysis.

Applebee's focus was on reading retardation but his principles have wider application.

It is clear from the nature of the data gathered in this study and the intervention hypotheses themselves that it would not be credible to offer simple solutions. Hypotheses based on single intervention and outcome variables, 'controlling for social class and I.Q.', brought valuable insights in earlier years but such methods are being superseded by the evident need for sophistication.

The essence of the approach to the analysis here is that virtually all measurable performance is the product of multiple variables; although it is impossible to measure all the contributors, the aim should be to build as many as can be identified into an hypothesised model of relationships and to test this model against the customary probability criteria.

The timidity of the statistical analyses applied to many intervention projects, in sharp contrast with the high quality of the measurement of the variables involved and the careful design of the research projects themselves, has been discussed by Barker (1978) in a paper on methodological issues in the statistical interpretation of intervention projects. The tendency is often to take a few individual parts of a complex model of relationships, carry out a fine-grained analysis of their associations and significance, and follow this with a study of such associations within subsets of social class, sex and other groupings. It is much less often that a global picture is examined, using the necessary advanced techniques. A comparable situation would be the difference between examining single trees and clusters of trees, and undertaking a study of the ecology of the whole forest. While both contribute to knowledge, the latter is too often ignored.

Two studies provides examples of these contrasting approaches. Both Neligan et al (1974) and Evans (1972) gathered large masses of data in their respective fields of research; each aimed at developing screening instruments, in the medical and educational fields respectively. While the Neligan study offers many suggestive findings on relationships between pairs and triplets of variables, it was the Evans study which employed the necessary advanced multivariate techniques enabling a model of relationships to be built; to that extent it was able to offer deeper insights into those relationships.

The fact that sophisticated techniques can offer more insights has its drawbacks, in that findings are more debatable and alternative interpretations can always be offered. The contentious issue of whether the American Follow Through programmes were 'successful' or not has been widely debated, with House et al (1978) and Anderson et al (1978) presenting diametrically opposed viewpoints. Without attempting to judge the differing statistical interpretations, the approach of House (1977/78) to objective experimentation and statistics is somewhat surprising. "The analysts thought objectivity was sufficient to ensure superiority and influence (in policy decisions based on the analyses).

More often it meant irrelevance. Objectivity sought to deal with interests by excluding them. What is needed is impartiality, which deals with interests by including and balancing them."

The implication of the House approach is that less tangible criteria (cultural, economic and other) should be allowed to displace the statistical conclusions, if the former are sufficiently weighty. It is an understandable but questionable approach since the new criteria are ill defined and essentially subjective. While they cannot be ignored, they can only be applied at the point where decisions have been made on the validity and the more general acceptability or otherwise of the statistical measures; they cannot be advanced, as House suggests at several points, as counters to the acceptance of the research findings.

In contrast to the implied rejection of statistical objectivity above, Goldstein(1976a) shows that the assumptions underlying a statistical design can have a powerful effect on the conclusions. He cites a set of data which can be used to 'prove' that either parental attitudes or school circumstances are more influential on child performance. It is the decision of whether to take as the 'utility function' all the performance scores in a continuum, or only those scores below a certain level, which point the conclusions in one or other direction. If one is planning an intervention programme for slow-learning or disadvantaged children, clearly the second utility function is more relevant.

Thus the issues surrounding analysis of the data are far from clearcut and are not resolved simply by taking necessary account of the complexity of the model.

There are several other guidelines in the approach to analysis in this study. Johnston (1972), writing as an econometrician, points out that the exclusion of relevant variables from a regression may be a very serious error, biasing the estimated coefficients; one should err on the side of including variables in an analysis rather than excluding them. He warns however that least squares does not automatically give the best linear unbiased estimates. It was these and many similar cautions in the statistical literature (referred to later in this chapter) which indicated that while regression models in the present analysis should be as complete as possible within the bounds of the data that could be collected, steps should also be taken to find promising theoretical alternatives to least squares and if necessary to develop these alternatives into practical methods of regression analysis.

The points made by Tukey, Johnston and Applebee above all emphasise the importance of collecting an adequate selection of variables to interpret the research model as fully as possible - but not including redundant variables, which can swamp the results - and setting up firm criteria for inclusion. Beyond

There is a considerable literature on the value of small samples. Cowles (1974) suggests  $N=35$  as a rule of thumb for researchers, and argues that large samples may encourage the acceptance of trivial results. McNemar (1969) reviews the interesting research findings on comparisons between very small samples (e.g. of 5 and 15) in a study of over 1,000 t-tests, most of which proved reliable despite the employment of non-normal distributions. Although Tversky and Kahneman (1971) and Gocka (1973) warn of the practical problems and cautions needed when relying on small samples, even studies based on  $N=1$  can yield useful results, according to authors such as Dukes (1970), Gottman (1973) and Leitenberg (1973).

None of these small sample studies have, however, attempted to take into account a multiplicity of variables. It is clear that the smaller the ratio of  $N$  to the number of variables in a model, the more open the results will be to misinterpretation due to random error in the variables or spurious relationships, influencing especially the estimation of regression coefficients in a linear model. Given such considerations, it was decided to aim at as large a sample as possible within the constraints of research capacity and the availability of child, parent and school samples. Other aspects of sample size, including the problem of attrition, are dealt with in section 4.20 and elsewhere.

The longitudinal nature of the research and the different conceptual groupings to which the variables belonged suggested that the basic regression method of analysis should be applied within a model of relationships to which path analysis could be applied. A possible alternative was multifactorial analysis of variance, for which a variety of powerful computer programmes have been developed (e.g. Multivar, GLIM and the latest versions of SPSS and BMD). While GLIM is probably the most statistically competent of these, relying as it does on maximum likelihood estimations, the fact that competing predictors have to be identified as binary or categorical entities in multifactorial analysis of variance models implies a loss of variance which reduces the explanatory power of these models. A further reason for choosing a regression model with continuous variables is that analysis of variance techniques cannot easily be applied within path analysis due to their overemphasis on orthogonal designs.

The work involved in developing, testing and applying the regression method and related path models meant that an early choice had to be made to rely on this single basic form of multivariate analysis, rather than comparing several forms of analysis as would have been possible with more than one research worker.

One final issue of general importance is the problem defined by Rosenthal (1976) and also by others as 'experimenter effect'. While generally applied to the danger of the experimenter influencing, often unconsciously, the behaviour of the subjects or even the test results themselves, it is an equally dangerous

possibility in the handling of statistical design and analyses. The misuse that can be made of stepwise and hierarchical regression in forcing in or excluding variables, as an example, is seldom recognised although this has clearly occurred in some major research studies of recent years; it is the complexity of the methods which makes possible the wrongful interpretation of analytical techniques that favour one or other result.

While design flaws are more obvious, there too are contending issues of practicality which sometimes force the employment of less acceptable models; the problem of control samples, dealt with elsewhere in this study, is a case in point.

The present chapter deals briefly with various statistical criteria, including bivariate and multivariate relationships, causal models, the assessment of reliability and validity, the use of disattenuation procedures, the significance concept and concepts of age and time. A major sub-section deals with regression analysis, in particular the development and validation of the V-ridge regression method in preference to ordinary least squares for data such as have been assembled in this study. Finally the issue of path analysis and the particular form used here are described in some detail.



### 5.10 Statistical options

The competing merits of laboratory and field experiments have frequently been discussed - in Kerlinger (1973), for example - and the arguments on both sides are fairly well known. It is generally accepted that, despite the imperfections of field research, its closeness to reality gives it greater validity than could be offered by most laboratory studies; thus the tentative conclusions reached from field studies can with some confidence be applied to other situations which are not necessarily an exact copy of the original research situation.

While the importance of field research is by now recognised, the statistical methods of dealing with the results are still, by and large, at a level that may be described as outdated and suited more to the highly controlled laboratory situation than to the multivariate reality of a field study. This shortcoming is not due to a lack of suitable methods. Statisticians have developed a vast armoury of multivariate techniques and new and more profound methods appear in the literature at intervals, but these are seldom applied to the areas of educational or developmental intervention research.

A review of a number of better known intervention studies supports this contention. Only brief reference will be made to each study, so as to focus on the issue of statistical technique.

Many of the studies which have gained fame on the basis of their findings have rested on little more than t-tests and comparisons of correlations across the experimental and control samples. Typical examples are the following: De Vries (1974), who undertook a study of language and cognitive development, using 42 experimental and 62 control children; Fowler (1978), who reports a study with a 'rolling sample' which totalled several hundred by the end of more than five years' intervention, comparing day-care and home-care children; Garber and Heber (1973), who compared 17 experimental and 18 control children in an intensive intervention programme, using an extremely disadvantaged sample; Gray and Klaus (1972), who relied on 88 children in four groups, one a distal control and the other three randomly allocated to different experimental/control treatments; Gutelius et al (1972), who based their study of cognitive stimulation in infancy on 46 experimental and 46 control children; Karnes et al (1968), who limited their pre-school study to 15 intervention and 15 control children; Wilkinson et al (1978, 1979), who undertook a study in a very deprived Glasgow estate containing about 500 children from 0 to 16 years; and Rosenthal and Jacobson (1968), who based their much disputed 'Pygmalion' study on 283 control and

70 intervention children, in approximately 15 classrooms. In most cases the samples were randomly allocated or matched on several criteria. These and many similar studies have relied on simple two and three-dimensional statistical models.

Four studies or groups of studies can be cited which have utilised more sophisticated methods of analysis. Hewison and Tizard (1980) report a study of parent 'coaching' of children, with 47 parents who coached and 53 who did not; regression analysis was used to assess the effectiveness of coaching, taking other factors into account. Kellaghan (1977) describes the evaluation of a pre-school intervention project in which outcomes were measured over a number of years; there were 90 experimental and 60 control children from a disadvantaged population and a further 60 controls from an advantaged population. The study made use of direct comparisons as well as discriminant function analyses to determine which variables were most affected by the programme. A summary of 12 major Headstart programmes, with sample numbers (experimental and control) varying from nearly 8,000 down to 26, is reviewed in Lazar et al (1977). The value of this study is that it looks at outcomes some 10 years after the intervention itself; a wide range of analyses were employed, ranging from sophisticated regression studies down to t-tests of differences. Another major American intervention project, Follow Through, was evaluated by Abt Associates (1977); some 20,000 children were involved in a wide variety of geographical and population groupings. Covariance analysis and t-tests were used. Both the methodology and the statistics were heavily criticised by House et al (1978) and defended by Anderson et al (1978) and others.

With the exception of the four groups of studies described in the last paragraph and a few other major projects, the statistics used to examine the results of most intervention programmes offer limited insights and in many cases may have failed to identify success owing to the crudeness of the analyses. There is often a ritualistic adherence to the view that provided one attempts to abide by the twelve criteria set out in the noted contribution by Campbell and Stanley (1963) - almost invariably cited as evidence of methodological propriety - there is no need for anything more sophisticated than a t-test, a few graphs to illustrate the widening difference over time along one or two variables, and perhaps a handful of correlations, all of which are adduced as evidence of significant post hoc sample differences or the failure to achieve such significance. The complexity of the situations in which intervention occurs, the near impossibility of ensuring completely equal control groups in volunteer field situations, and the multiplicity of contributors to change are ignored in favour of a limited empirical approach in which p of .05 or .01 is the final 'proof' of the success or failure of the experiment.

The economic cost of the effort put into bringing about change is hardly ever considered, with the result that great differences such as those secured in the Garber and Heber (ibid) Milwaukee study achieve a popular importance which bears no relation to the immense resources of personnel and funding put into a small number of children and their parents.

Two important studies, a quarter century apart, are among the few statistically oriented reviews which have examined the quality of intervention projects. McNemar (1945) examined what were at that time regarded as important studies of the effects of intervention in orphanage samples and showed that the well-meaning practice of control sample replacement during the course of the research led to spuriously good comparisons, favouring the experimental sample. On the other hand Campbell and Erlebacher (1970) reviewed the application of covariance analysis in regression models, as used in Headstart and other intervention studies, and concluded that a misapplication of this technique had led to the impression that compensatory education programmes were harmful.

While those and similar studies have pointed to particular shortcomings, the principle of applying t-tests or correlational analysis to intervention in a field research situation has not usually been challenged, perhaps because of claims that the experimental and control samples were well matched - though this is rarely possible to achieve on more than a few variables. In consequence it has been possible to ascribe importance and 'significance' to a host of competing intervention variables predicting the same few outcomes. Multivariate analysis has yet to be accepted as a routine procedure in this type of research.

One group of authors who discuss the issue of complexity are Namboodiri, Carter and Blalock (1975). They refer to the criticism of the use of advanced analytical techniques, criticism based on the claim that such techniques outstrip the quality of ones data and theoretical knowledge. The authors point out that if one fails to apply rigid controls through laboratory type experiments, then the analyses should in fact take the resulting complexity into account. They state that the overwhelming majority of quantitative studies in sociology, political science, psychology and economics make use of data analytic procedures that are far too simplistic for what is commonly assumed to be true of the real world. There is a sense in which it may be argued that the poorer the quality of ones data, the more complex should be the analysis in order to correct for such deficiencies.

The authors also plead for a higher level of statistical literacy on the grounds that the average level within a discipline is likely to determine what studies get read and become part of the body of literature in the field. If readers are not capable of understanding more advanced techniques, researchers are likely to utilise simple techniques from which many important insights may

be lost.

Namboordiri et al point out that if one was faced with five sources of error it would hardly make sense to refuse to correct one source (e.g. by using sophisticated analyses) on the grounds that the other sources could not also be eliminated. That would imply an all or nothing approach to research. The reality is that most errors are not additive or cumulative by nature. Thus if one does not have a perfectly random sample this does not mean that the biases need be serious ones. If an error distribution is non-normal, this does not imply the avoidance of tests requiring the normality assumption, as the tests may be very robust or insensitive to the violation of a particular assumption. Furthermore, if one does not have a perfect measure of  $x$ , it does not indicate that one should have resort to treating the variable as a dichotomy. If the 'dependent' variable in an equation is thought to have some very minor effect on one of the independent variables, this does not automatically rule out least squares analysis. Only if all these assumptions are violated to a major degree can one be certain of difficulty. Implicit assumptions should however be brought into the open and data analyses should match the complexity of reality. On the other hand, just enough complexity should be added to a model to cope with that reality.

The relatively new multivariate approaches to analysis by these and other authors suggest that the traditional application of Occam's razor might be modified to state that there is first a basic assumption of complexity, with the razor principle applied only at the point at which competing models each ensure a full consideration of this complexity within the field research situation; only then should the choice be made in favour of the more parsimonious of two models.

If the need for some complexity in analysis is accepted, it would also seem necessary not to forfeit any of the variance which might reside in the data - however burdened they might be with error - by rounding off or grouping scores into categories. Unnecessary truncation of data can only reduce the likelihood of accepting research hypotheses. Thus, for example, the key intervention variable of meetings attended in the present study is used as such, with adjustments for meetings partially attended, rather than treating the intervention as a dichotomous variable. At no point in the analysis have the data been collapsed on the grounds of numerical or statistical convenience; with access to computers such short cuts can no longer be justified.

### 5.101 Descriptive analyses

Rudimentary summaries of the data, such as the first four moments of each variable (mean, standard deviation, skewness and flatness or peakedness, the latter leading to the calculation of kurtosis), are invariably the starting point for a fuller study of research findings. Similar summaries can be carried out on sub-groups of the main samples at an early stage, especially to indicate whether the sub-groups differ from each other or from the main sample parameters. Histograms are a particularly valuable adjunct, as they present a visual picture of distributions.

Beyond this stage are the basic statistical measures of relationships and differences, such as bivariate correlations and t-tests. Once this point is reached other criteria come into play for assessing the probability of likelihood of a particular parameter. The practice of 'data dredging', the atheoretical inspection of large numbers of relations or differences in the expectation of finding something useful, is one that has to be approached most cautiously.

On the one hand the probability of obtaining any one of these bivariate relationships or differences by pure chance is given by the ratio

$$1 : \sum_{r=0}^{r=m-1} (1 - p)^r$$

where  $p$  is the probability of a finding in the face of the null hypothesis  
and  $m$  is the number of relationships examined

On the other hand an examination of such basic parameters may well point to an unexpected relationship or difference whose meaningfulness and probability require further and more reliable assessment

The above measures clearly refer to parametric statistics. It is doubtful whether non-parametric statistics have any important place in most multivariate situations. Bock (1975) points out that despite the usefulness of such statistics in special situations, they are too narrowly involved with tests of null hypotheses to provide by themselves a sufficiently comprehensive account of the complex data typical of behavioural studies.

### 5.102 Interpretive analyses

It is at the level of analytical interpretation that multivariate data can offer rewarding insights into hypothesised models of relationships.

A fundamental difference of approach to interpretation has been developing among statisticians over many decades. Today it is usually described in terms of the mainstream frequentist approach and the minority Bayesian approach. Hays (1973) presents a brief summary chapter on Bayesian concepts, while Theil

(1971) offers a mathematical exposition. However it is Edwards (1972) who discusses the Bayesian position in great detail and argues that it cannot meet the challenge offered by the maximum likelihood method (which is essentially frequentist).

The essence of the Bayesian position is that prior probabilities and a certain involvement of subjective experiential judgement are justified as statistical input when assessing the probability of outcome measurements. A major exponent of modern Bayesian thinking is Lindley (1965).

The purpose of referring briefly to this debate is to acknowledge that Bayesian analysis does offer an alternative method, though inadequately developed for handling multivariate data sets, in which the probability of a set of research findings is not adjudged as arising from a theoretically unknown 'black box' but rather from a known set of prior probabilities.

Within the mainstream approach the possibility of using maximum likelihood methods (MLM) in this study was considered at an early stage, especially as important computer analytical packages such as GLIM are based on MLM algorithms. The decision was taken not to use these methods since they require a number of fairly stringent assumptions which are unlikely to be met within the present set of data. Klein (1974) points out that one requires strong a priori specifications of the model being employed (e.g. regarding the variables included and excluded) before employing MLM, otherwise estimates throughout the system can be disturbed. Theil (1971, pp.609-610) refers to several problems in applying MLM to regression equations.

The possibility of interpreting the research findings in terms of a Markov process model was not seriously considered because of the limited scope of using this model at present. But the approach does merit some mention because Markov chains are essentially a model for interpreting discrete rather than continuous changes in behaviour. Thus a change in a parent's marital state is likely to represent a major and sudden change of the behavioural environment surrounding a child, compared with the continuum of change represented by a parent's gradual acquisition of parenting skills. While the latter can form part of a regression model, in which parenting skills are but one of a variety of continuous or categorical variables, the effects of sudden environmental changes need interpretation within a Markov matrix of probabilities.

Ultimately a unified theory which succeeds in marrying the potential of regression and Markov models, possibly with separate equations within each Markov cell, may afford more sensitive insights into behavioural patterns than have yet appeared in the literature.

Francis (1980), for example, suggests that the intermixture of gradual and sudden changes in a model may be handled in terms of the recently developed

catastrophe theory. One of the variables in a situation may increase to a point where other variables quite suddenly shift on to a different plane. She also argues that within a complex system of interacting variables, categorical and continuous, the relevant mathematical models for data analysis might include the mathematics of rates of change in multi-dimensional space rather than the mathematics of sampling distributions.

Having looked briefly at several fundamentally differing approaches to the analysis of data, one can focus on the linear model which, in terms of its wide scope, its possibilities for theoretical and practical interpretation and its sophistication appears to be the most suitable analytical model for this research. Whether aimed at testing hypotheses for the significance of effects or group differences (analysis of variance, discriminant analysis, contingency tables, etc.) or assessing and testing a predictive model (regression analysis), the model relies on interpreting variance after taking account of error.

Other sections of this chapter deal with the various forms of descriptive analysis and the particular linear model used here. The section below looks briefly at four techniques which were considered but rejected on theoretical grounds.

### 5.103 Particular techniques and their problems

Some of the statistical techniques which were considered prior to the design and analysis of this study included the following.

Change scores. Although the analysis of change for certain specific purposes can be justified, such as the assessment of the correlates of test changes over time in a survey population (Fogelman and Goldstein, 1976, for example), or the examination of changes in a variable assumed to remain fairly constant, such as the study by Hindley and Owen (1978) of individual changes in IQ over a long period, in general the use of change scores is regarded as a problematical procedure, especially when assessing change resulting from an intervention process.

There are many weighty supporters of this view, including Cohen and Cohen (1975), Campbell and Erlebacher (1970) and Harris, C. (1963). Perhaps the most convincing criticisms of the use of change scores for assessment come from Lord (1956, 1958, 1963) and Cronbach and Furby (1970). They point to the major part played by error in the composition of change or difference scores. In general, regression techniques are preferable for handling or interpreting changes in scores; if necessary the initial scores on a dependent variable can be used as an independent contributor to the final scores.

Outliers. A considerable number of studies have been devoted to methods of handling outliers - scores which are so extreme that they may well be erroneous or may be representative of some other population than the one whose characteristics are being examined. In statistical terms outliers are undoubtedly an embarrassment. An outlier 3 standard deviations from the sample mean will have a second moment 'weight' nine times that of cases one s.d. away from the mean; in relatively small samples the effect of one or two outliers can be such that a significant effect may be converted to non-significance, or vice versa. Correlations and other parameters of the sample may also be seriously distorted.

Tukey (1962) suggests methods for 'trimming' what he calls spotty data - by eliminating all data above and below certain levels, or by Winsorising, that is, replacing outliers with the nearest value of a non-suspect observation. Anscombe and Tukey (1963) argue that possibly the main justification for calculating residuals is to detect the presence of outliers; the authors suggest various techniques for deciding on rejection or retention. Ellenberg (1976) points out that outliers may result not only from a gross reading error or other obvious mistake but also from incorrect parameter specification or non-normality of the error term. An interesting contribution is offered by De Finetti (1961), who contends that a Bayesian analysis overcomes the problem of outliers by drastically reducing their influence on the final distribution, relying on the initial parameters to bring about this result.

A variety of sophisticated jack-knife methods have been developed to cope with outliers according to formal statistical procedures rather than the ad hoc decisions of the analyst. Sharot (1976) offers one of many reviews of the effects of these procedures on reducing the bias of the estimator.

While the respectability of such methods has to be acknowledged, the reality of tampering with seemingly 'unacceptable' scores brings with it major questions on the validity of the results achieved after dismissing or altering certain data. Although formal jack-knife procedures may be invoked in statistical support of a research decision to question the use of a particular score value, the procedures would have to be seen (and explicitly justified) as deriving from the analyst's initiative rather than from a blind statistical process, unless the same jack-knife was being routinely applied to every data set in that study.

Matched groups. The use of matched groups in research experiments had a long and modestly successful history in the days when children's (or adult's) behaviour levels were seen as the outcome of a few predictors. Experimental and control groups were matched on several key variables and the remaining variables served as the experimental intervention. With the growing realisation that a large number of variables contribute to almost every major outcome in education and



psychology, the use of matched groups has declined considerably. It is seen to be almost impossible to have a large enough sample from which to select two groups matched on all the variables thought relevant to some outcome.

Burroughs (1971) presents a trenchant criticism of the principle of matching, cited inter alia the problem of the decreasing randomness of a population as individuals are removed from it in the matching process.

A more fundamental criticism is offered by Campbell and Erlebacher (1970), who point out that matching is an ubiquitous error commonly found in quasi-experimental situations, in an attempt to overcome pretest inequality between groups. The authors point out that selection of a subset from a group (in order to match it with a subset from another group) is likely to involve the phenomenon of regression to the mean, since individuals within the two matched groups will not have been chosen randomly from those groups. The use of multiple matching variables may reduce the regression artefact but it will not remove it. Even matching by means of dichotomous variables has a bias, since all matching variables are imperfect indicators of the underlying concepts on which matching is sought.

Recent attempts to undertake matching, such as in the study by Woodhead (1976a), showed the difficulties involved in carrying out this procedure properly and found that it was virtually impossible to establish an equivalent control group at any point in what was generally an ambitious and well monitored study. This experience, and the conceptual problems referred to earlier, make it evident that matching did not offer a viable approach to design or analysis in the present study.

Control and covariance. The methodological problems associated with establishing a control sample in the present study have already been discussed (section 4.226). The statistical issues surrounding the concept of control need a brief review. In essence a control sample is seen as differing from an experimental sample only in terms of a single treatment variable. This is rarely achieved in practice because of the difficulty of establishing a completely random allocation of cases into experimental and control samples.

Kish (1975) discusses the types of variables which arise in a research situation. The goal of the researcher is to put as many disturbance variables as possible into the category of control variables, to avoid their being confounded with the explanatory variables. While the division of the samples into sub-classes based on specific levels of a disturbance variable offers one possibility for overcoming the presence of such a variable, the method often applied in practice is to adjust outcome scores by covariance, that is, by computing the covariance of the dependent variable with the disturbance variable. The former variable is then adjusted to take account of the regression effect of the latter.

Campbell and Erlebacher (ibid), Gourlay (1953), and Cohen and Cohen (1975), among others, discuss the serious difficulties attendant upon the covariance method. Gourlay sets out the stringent conditions required for use of the technique, while Cohen and Cohen show how error in a partialled variable can have serious consequences by distorting the analysis.

It is Campbell and Erlebacher's study which can be regarded as a definitive statement on the basic problems with analysis of covariance. They demonstrate how the failure to correct for the fallibility of the partialled variable could well have made the Headstart intervention programme appear ineffective. The authors criticism is directly specifically at the research undertaken by Westinghouse/Ohio (Cicirelli et al, 1969), a study which is still cited today as 'evidence' of the failure of Headstart.

While these and many other authors discuss the use and shortcomings of covariance methods in terms of the fallibility or characteristics of the control variable or variable set, few focus on a more fundamental conceptual problem with covariance. It is in the nature of behavioural research that most predictor variables are intercorrelated, sometimes highly, and consequently contribute both unique and shared variance to a dependent variable.

The essence of any control procedure is that the controlling variable removes from the outcome variable the whole of the variance with which the controlling variable is associated - unique variance as well as that part shared with other predictors. This occurs even when a sample is divided into sub-classes based on specific levels of a disturbance variable. The procedure puts a grave limitation on the possibility of showing intervention effects, since the probability of an effect is a function of both the absolute size of the effect in competition with and alongside other effects, and its standard error derived from the whole matrix of effects. The unique variance of an intervention effect

- the rump left after removal of the covariate's unique and shared variances
- defines only part of the importance of an effect.

A way out of this dilemma - one followed in the present research - is to treat all possible predictors as contributors to outcome variance and deal with the problems of multicollinearity by the use of a variety of techniques and checks, as described later in this section. It is hoped to deal with the limitations of the study's control sample in terms of a macro analysis rather than attempting to take account of differences (whether of means or contributions to variance) by covariance methods.

The remainder of this section reviews some of the broader statistical methods which are of particular value for the evaluation of research programmes in education and psychology, and the reasons for the choice of certain methods of analysis in this study.

### 5.11 Bivariate relationships

Bivariate relationships are somewhat unusual statistical parameters. They are basic building blocks for some major analytical methods, such as regression and canonical correlation, and they serve as a crude indicator of whether relationships exist or not. In themselves, however, they are highly fallible and the attempt to use the building blocks to define the building has been heavily criticised.

Tukey (1954) presents a classical discussion of the issue, a discussion which is still relevant today. He sees the correlation coefficient as tangential to data analysis; it shows what the effects could be of small changes in one or other variable but it does not embrace a wide range of values or even of samples. It can never rise to be a functional measure. Even the use of correlation coefficients in deriving a model of path analysis, as Wright did, is a tangential method; the coefficients do not become functional until they are translated and interpreted as regression coefficients. The correlation coefficient itself can be justified only in limited areas of analysis, such as in genetic work.

The worst abuse of the correlation coefficient arises when it is used in an attempt to define causation. Simon (1971) shows with a simple three-variable model the considerable variety of possibilities (including reciprocal relationships) which may be concealed within an ostensible bivariate relationship between two of the variables. The difficulty with Simon's model is that its particularistic approach suggests that by using his criteria for judging the direction and directness of a relationship within a three-variable model, one can reach some conclusion about causation. This is not so, since even the simplest of relationships would need to be tested within the framework of a relatively large number of variables.

A different approach is that taken by Cronbach (1970), who argues that Yule's method of partial correlations (the precursor of modern factor analysis) enabled the isolation of variables which could not be isolated experimentally. He considers that the experimentalist and correlational approaches have much to contribute to each other. Like Simon, Cronbach does not go on to examine his theories within a wider matrix of relationships, where the use of correlations for anything other than building blocks becomes untenable.

Another defence of correlations is offered by Havlicek and Peterson (1977), who used Monte Carlo procedures to test many thousands of samples ranging down from 60 to 5 cases; they found that the Pearson coefficient is insensitive to extreme violations of the basic assumptions of normality and of the type of measurement scale (for example, non-equality of intervals). This is useful evidence to support the contention of Kerlinger and Pedhazur (1973) that regression

itself is also extremely robust when its assumptions are not met. It is the strength of the building blocks in the face of deviations from basic assumptions that contribute to the robustness of regressions based on those primary correlations.

A major difficulty with the use of correlation coefficients for anything other than building blocks is that the doubts that must exist about what may be a spurious relationship (due to the presence of one or more intervening variable) virtually negates the significance test as a cautionary check on the strength of association. The smallness of a probability can only serve as contributory evidence for the directness of a relationship, on the grounds that a mediated relationship is unlikely to be that strong.

A somewhat unusual defence of correlations is offered by Kendall and Stuart (1963) in their well known statistical text. They consider that the reaction to the dangers cited by Yule in 1923 (on the evidence that high correlations are not necessary causal) has gone too far and correlational evidence is now very unfashionable among statisticians. These authors consider that in large areas of the social sciences and psychology the patterns of causation are not yet sufficiently well understood for correlational analysis to be replaced by more specifically 'structural' statistical methods. Even given the limitations of regression and path analysis in 1963, compared to today's models, it is a rather surprising statement. The fact that structural methods are sometimes abused, or that causal models are derived without sufficient confirmatory evidence, is not a convincing argument for a return to reliance on a correlational statistic whose severe limitations were described nearly half a century ago.

Blalock (1964) deals with yet another flaw in correlational methodology - the considerable fluctuations that can occur when cases are grouped. With appropriate grouping of cases it is possible to elevate correlations well above the levels of the 'raw' correlation based on the individual cases. Even when there are good grounds for grouping, Blalock's evidence and the reduced size of group samples point to the likelihood of increased error.

Correlational analysis does have its uses, nevertheless, as was suggested at the start of this sub-section. It is a quickly derived and easily understood basic statistic with which to start off an examination of any model of relationships. It can be used expeditiously within sub-samples for a visual examination of possible relationships before embarking on macro analyses. It can serve equally easily, though not without some doubts, as a criterion measure for examining the residuals derived from major analyses such as regression.

Its most important function is, however, as the creator of the building blocks already referred to. When used in this form correlations have a powerful

role as the foundation of regression analysis; the shortcomings of the individual blocks are usually (though not always) sharply reduced in a situation where it is the inverse of a whole matrix of relationships which determines the regression coefficients, establishing clearly the dominance of the multivariate explanation for an outcome rather than the oversimplified interpretation offered by ordinary bivariate relationships.

At this point mention can also be made of canonical correlation. As a specifically multivariate technique - relating multiple variable sets to each other - it is in some ways an advance on multiple regression; it creates its own 'principal component' variables and may be a useful approach where multiple predictors and multiple criteria exist. However it has no provision for sequentiality as occurs in path models, for it simply maximises the multiple correlation between two sets of variables. To that extent it suffers, at a highly sophisticated level, from some of the shortcomings of correlational analysis referred to above, although a recent study has suggested wider possibilities for this technique.

### 5.12 The multivariate approach

Given the problems which arise in attempting to interpret data in terms of bivariate relationships it is evident that the possibilities of a multivariate approach were strongly considered. The variety of sources of variance in the present research was already suggesting the need for a sophisticated model of relationships. There was also a wealth of evidence and argument from statisticians on the desirability of a multivariate approach to data analysis.

Kendall (1957) points out that any set of variates measured on subjects are dependent among themselves; one or more cannot be split off and considered by itself. Cronbach (1976) puts this view more abruptly: "People are fed up with reports that summarise student attainment in a single index and then conclude that no differences between treatments can be detected..... It takes multivariate measures to disclose patterns of achievement." As McCall (1970) points out, responses in nature do not occur in isolation and analysis should take account of this rather than ignoring the interrelationships among several variables. Comments by Namboordiri et al (1975), in support of a more global view of data analysis, have already been cited. Despite the practice by many researchers of focusing on over-simplified models of relationships the statisticians are nearly unanimous in favouring more complex interpretations.

The crucial issue to be decided upon was the exact nature of the multivariate approach to be adopted in the present study. There was little problem about accepting the theoretical position that, with the exception of a few categorical variables such as sex, the continuous variables and the model as a whole fitted reasonably closely to a multivariate normal distribution, despite the presence of a few highly skewed independent variables. The major problems in applying a multivariate model to non-parametric data (which have been tackled by workers such as Boyle (1970) and Kim (1975) ), precluded any treatment of the data other than as part of a parametric situation.

Among the various considerations born in mind when deciding on the form of analysis were the opportunities provided by a modestly large N (given the nature of the research), the fundamental question of parameter sensitivity within any one model, and the particular difficulty resulting from possible reciprocal causation between variables.

Both Werts et al (1974) and Green (1977) refer to the most basic problem facing the analyst of multivariate data, namely how adequate or conclusive is one particular model compared to some other possible model of relationships. What Werts et al describe as overidentification of a model (when the variety of interlocking equations relating variables within a matrix of relationships is more than adequate to derive a single solution to those equations) implies that

one can select a subset of equations to yield one particular solution rather than another solution. It is extremely difficult to show which model best simulates reality, since several models may produce convincing levels of significance for the derived parameters. Green describes this problem as one of parameter sensitivity; a sensitive set of parameters will offer a minimal solution to a model; a less sensitive set will be more likely to yield a chance fit to the data; again the difficulty is one of identifying the more credible alternative. Werts et al suggest that once the possible range of models has been derived it will be necessary to determine what further data is needed to discriminate between the models. The particular difficulty of reciprocal causation is discussed by Namboordiri et al (ibid) among others. The problem here is that the existence of reciprocal causation between pairs of variables makes the derivation of a model increasingly difficult as the number of such mutual influences grows.

Fundamental problems such as those discussed above do not point in the direction of any particular multivariate approach; some approaches may serve to conceal the problems in so far as they are not clearly identifiable in the derived solution or model, but no approach can eliminate the problems.

Finn (1974) sees multivariate techniques as comprising two related methodologies. The one is concerned with the discovery of an underlying structure of response data - essentially some form of component or factor analysis. The other is concerned with the contribution of structured variables to the variation between individuals or groups. The merits and demerits of each methodology are discussed briefly in the next two sub-sections.

#### 5.121 The component and factor hypotheses

As Cooley and Lohnes (1971) point out, multivariate procedures are often concerned with reducing the original test space to the minimum number of dimensions needed to describe the information in the original observations. This in essence is the concept underlying the development of both component analysis and factor analysis. While sharing this goal of reducing the test space, the hypotheses and mathematics underlying the two forms of analysis differ sharply and the latent or underlying variables derived in each form of analysis have very different characteristics.

Principal component analysis is a straightforward technique, with relatively few variants, by which the total variance within a set of variables is reproduced by a set of orthogonal variables equal in number to the original set of variables. Since the method of deriving these variables maximises the variance contained within each successive component, a smaller number of components are required to

reproduce most of the variance in the original set. Some variants of principal component analysis apply the method to a subset of variables and the major components are then used as covariates in a model within which the remaining variables are used in their original form. Regression analysis can also be carried out using the most important components rather than attempting to regress the outcome variable on all the raw predictors; this assumes, however, that the nature of these components can be clearly identified if there is to be any adequate interpretation of the model. The regression itself is straightforward as the orthogonality of the components eliminates all problems of multicollinearity.

Factor analysis, in contrast to component analysis, summarises the inter-correlations within a matrix in such a way that the derived set of common and unique factors reproduce maximally the original matrix. Harman (1967), whose presentation of factor analysis is one of the major contributions to this field, shows that this form of analysis has a vast range of possibilities and variants, orthogonal and otherwise, with a selection of rotational solutions as additional options. Apart from the differing methods of derivation, the most important conceptual difference between component and factor analysis is that the latter makes what seems understandable provision for a 'unique' factor for each variable (representing what is not shared with other variables), as well as a number of shared factors covering the sources of common variance between variables. Factor analysis also has variants which enable the regression of a dependent variable on the derived factors; Lawley and Maxwell (1973) have developed an interesting maximum likelihood form of factor analysis, with a chi-squared criterion used to limit the number of factors; Schönemann and Steiger (1976) present a regression component analysis which is not indeterminate and offers a solution rather different from that of factor analysis.

There are two fundamental criticisms of the component or factor approach, one conceptual and the other based on a range of statistical challenges. Conceptually the most difficult feature of either of these models is the assumption that the division of the overall variance or the re-arrangement of the correlational structure between the variables will produce a set of components or factors within intrinsic credibility as representing certain latent variables. While component analysis presents the rather simple solution of maximising the variance of each successive component, factor analysis offers such a vast range of statistical variants and alternative solutions that almost any explanation can be fitted to these solutions. Cureton (1939) makes a bitter attack on intelligence and personality specialists for applying "higher mathematics to wishful thinking". The solutions themselves are far from unique, adding to the difficulty of proof.

In statistical terms there are many criticisms. Even component analysis is not unique, as the solution alters if the scale of any of the variables is altered;



furthermore, as Maxwell (1977) points out, the variance extracted includes all the error variances and makes it indistinguishable from the true variance within the variables. Van de Geer (1971) states that the solution of a principal components model is completely changed if one of the variables is added or removed; within factor analysis one starts with a mathematical model dependent on certain assumptions, but if the model does not work then one decides post hoc on an amended model. The criteria for acceptance or rejection of a solution are extremely loose, particularly when solutions can be rotated to obtain a more credible set of factors according to a wide range of criteria.

These and many other criticisms do not imply that the search for underlying components or factors is valueless; solutions are often in accord with 'common sense' and the loadings of a major factor on a set of variables (particularly if a marker variable is included) may seem quite credible in relation to a particular hypothesis, given the uncertainty of any one solution. Nevertheless the use of these methods, or even their regression variants, did not seem appropriate for the present study where some measure of finality is being sought on the question of whether the parent programme had been effective or not.

#### 5.122 The interpretation of variance

One of the most important statistical tools available today is the division of the variance of a variable between other variables thought to relate to the first variable in some way. A sensitive examination of shared variances may tell us a great deal about hypothesised relationships. Today this is an accepted notion and yet it is only in recent decades that the idea of 'explaining' variances as a means of interpreting a model has come to prominence in statistical analysis. Two major forms of analysis within this paradigm are analysis of variance - probably the most widely used technique in advanced educational and psychological research today - and regression analysis.

The origins of the statistical concepts underlying these methods go back to Legendre, Laplace and Gauss, who early last century made separate contributions to the development of the least squares method (fitting a line to a series of points by minimising the squared deviations from that line); in the same century Galton developed the concept of regression, and body was added to the least squares statistical model by Markov early this century. Other historical milestones were Pearson's development of the correlation coefficient at the start of this century, followed by Ronald Fisher's development of analysis of variance in the 1920's.

The relevance of these few historical references is that while the development of regression concepts preceded that of analysis of variance, it may well

have been the seemingly greater certainty and more clearly specified rules for determining different probability criteria which led to Anova (analysis of variance) becoming the principal technique for many analysts.

The regression model, its limitations and a new variant of the regression model are detailed in sections 5.20 to 5.23. At this point it is only necessary to state briefly why regression has been chosen in preference to Anova for the present study.

Anova offers many useful statistics, including clear criteria for deciding whether two or more samples differ significantly and whether certain variable 'effects' are significant in relation to an outcome variable, even in the presence of other predictor variables. There are a great many forms of Anova and computer programmes enable highly complex models to be analysed. Regression, on the other hand, offers fewer decision criteria and the presence and problems of error are more evident than they are with Anova - although in fact errors in variables and errors in specifying the hypothesised model of relationships can affect both methods equally seriously.

The reasons why regression was preferred in this study are the following:

- a. Regression can handle a large number of competing variables.
- b. The planned derivation of a path analysis model is better served by regression equations.
- c. Continuous independent variables have to be transformed into blocks or categories for use in Anova, with a consequent and sometimes serious loss of variance (except in the case of Ancova, which has its own major limitations).
- d. The basic approach of Anova tends to be that of presenting the significance of differences or effects across a small number of groups or levels of a variable; in contrast regression presents the significance and relative 'weight' of each independent variable in predicting the dependent variable. Interpretation in this study thus appears to be better served by the latter method.
- e. Regression does not face the problems of unequal cell sizes which complicates the use of Anova.
- f. The statistical shortcomings of regression, which are discussed at length in section 5.21, are evident and clearly identifiable, contributing thus to methods by which these limitations can be reduced. In contrast the statistical shortcomings of Anova are less easily disentangled from a more complex presentation of outcomes.

A final...../

A final reason is that the development in this study of a modified ridge regression, termed V-ridge, has brought about a sharp reduction in the level of error in the regression coefficients, as judged by cross-validation trials.

### 5.13 Causal models

There are two major issues which arise in the presentation of a causal model within a research study. The first is the philosophical question of whether causality can ever be propounded as a scientific proposition, and the second is the statistical question of whether a particular model can ever be shown to be the true model.

There may be less difficulty with the first question than with the second. Using the criterion of falsifiability propounded by Popper (1976), it is always possible that the discovery of new variables or unknown relationships between existing variables may disprove a particular model; causality in these terms can thus be seen as a scientific hypothesis. On the other hand there is a long tradition in certain branches of English philosophy which goes counter to attempts to establish any causal models.

Thomas Hobbes (1588-1679) regarded senses as the source of all knowledge, with memory and imagination being decaying sense impressions held together by a train of thoughts based on contiguity and other associations. John Locke and David Berkeley carried these ideas further and it was David Hume (1711-1776) whose development of the 'Laws of Association' and rejection of the conception of 'mind' had a major influence on succeeding generations of thinkers in England. This essentially associationist approach, rejecting deductive models, had in parallel a healthy empiricism in which the feasibility of an idea could be tested by its practical outcomes without carrying a heavy burden of theory alongside.

In contrast was the Continental development of nativism, with Emmanuel Kant (1727-1804) as its major exponent. He argued for the imposition of categories on experience to make it intelligible, with the mind actively developing explanations on the basis of what is learned from the senses. He is regarded as a major influence on the development of Piagetian psychology.

It would be an oversimplification to attempt to define these approaches in totally contrasting terms, but it is reasonable to argue that the early English school influenced the development of analytical strategies which examine and test associations and work inductively towards the acceptance or rejection of hypotheses about those associations; on the other hand the early Continental school has led analysts towards the development of deductive causal models within which particular hypotheses can be examined.

A particular strength of the former has been the development in the late 19th and 20th centuries of sophisticated models of probability with which hypotheses can be tested. The strength of the latter lies in the wider perspectives afforded by causal models which, while less open to the same extensive range of proba-

bility tests, do nevertheless challenge the analyst to develop yet more comprehensive explanations. In many ways the strengths of each approach are the weaknesses of the other.

The choice of a research design clearly involves these philosophical considerations, as well as the range of analytical and computing techniques known or available to the analyst and the specific questions being asked or hypotheses being examined within the research theme. The theoretical questions posed in the discussion of this thesis up to the present section make it evident that a causal model is more likely to offer the range of testable hypotheses and macro explanations than would a narrowing of the research perspective to a more fine-grained and more accurate but less exploratory analysis of the research data. Within the limits of designing and carrying out a single study, with its lengthy analyses, it would have been difficult to examine the data competently according to both models.

There are as many doubts and unresolved issues concerning causal models as there are around models based on the testing of associations.

The statistical problem of establishing the truth or validity of a causal model is as difficult as the philosophical issue surrounding causality itself. It is virtually impossible to show that a particular model offers the only statistical explanation for an effect or outcome. On the other hand the use of logic and the linking together of psychological, educational and other concepts can offer a tenable model within the present knowledge in these fields and, in the absence of a better or statistically more credible explanation, it is reasonable to put forward a particular model as the best which can be derived in terms of present knowledge.

Blalock (1964, 1971) discusses causal models in considerable depth. He points to the question of whether X can be seen as a necessary and sufficient cause of Y. This brings with it an examination of alternative explanations of Y and the need to study how controllable are other variables which might contribute to Y. He examines the complex issue of spurious relations, as does Simon (1971) in one of Blalock's volumes, and he and his fellow contributors look at many other problems which arise within this area. It is worth noting that Blalock's focus is directed at non-experimental models, which by definition are less open to the testing of the hypotheses underlying those models.

Perhaps the biggest doubt about causal models arises from the fact that virtually every outcome variable has a multivariate set of predictors. This means firstly that even when all known predictors have been measured it is highly likely that there are yet more unmeasured variables which should be taken into the model; secondly it leads to a level of under-prediction in that only part of

the outcome variance can be accounted for - often only a limited part, such as well under half of the total variance. This is particularly the case in the social sciences, where most variables are seen as 'soft' and difficult to define or measure accurately. Even if disattenuation takes account of unreliability of measurement one is still left with the puzzling question of why so much variance cannot be accounted for; there are several ways of explaining this failure.

The most obvious is that important but hard to measure contributors have been omitted from the model. For example, a range of environmental variables on a child and his or her home environment cannot include the more intangible genetic contributors to performance (even when pre-test performance serves as some indicator of genetic contributions to post-test performance). A more difficult problem is that the ways in which the independent contributors act to produce the outcome variable may be more complex than can be understood at present; there may be unknown intervening variables which serve to modify the effects of the measured independent variables. The personalities and relationships involved with a particular child and his adult environment are not easily assessed, and yet these may have a marked effect on the cognitive and motivational skills which the child brings to a learning situation.

Interaction and temporal effects are also problematical in setting up a causal model. It is easy to create a simple interaction variable from two variables hypothesised to have a strong joint effect on an outcome variable. But as Hindley (1977) points out, there can be uncertainty as to whether a variable such as social reinforcement has a primary or secondary effect on a child; in other words, is it a direct contributor or a mediating variable for motivational variables within the child? Many school variables are hard to define, such as the stability of an environment; measurement of identifiable features of the school environment at one point in time may not reveal the regularity or reliability of those features.

Reciprocal causation is a particularly difficult matter to handle, conceptually and statistically. It is not always easy to determine the direction of causality of the interrelated variables. If they are seen as reciprocally causative it is necessary to use special techniques. As Namboordiri et al (1975) show, techniques such as the use of simultaneous equations can take this into account and researchers need not retreat into "mysticism or defeatism" in the face of such difficulties. Even so there are unresolved difficulties, particularly in regard to probability criteria for the acceptance or rejection of parts of a complex model.

Another problem which is more likely to arise in a complex model than in a simple two or three-variable experimental design is the possibility of correlated

error within a large number of variables. In other words, errors of measurement may, for a variety of reasons, be correlated across two or more of the variables in the model, threatening the reliability of the parameters by which the model will eventually be judged. It is probable that there is some very minor level of correlated error even within a simple design, but the statistical blurring from this source of error becomes larger and of greater concern as the model is expanded. As Goldberg (1971) suggests, the assumption of uncorrelated error is almost always invalid, but he considers that science proceeds in spite of this invalidity, while continuing its attempts to cope with it.

Decisions about the use of standardised or unstandardised variables are a further source of difficulty, since a complex causal model, sensitive as it is to the interrelationship between its constituent parts, is altered to some modest degree by the decision of whether to standardise the variables or not. There are many arguments on both sides and the issue will be referred to again, under section 5.30.

Given the many and weighty problems which arise with a causal model it may be asked why then proceed with an analysis based on that model? The answer is that the new perspectives afforded by causal models, the openings into alternative explanations, and the prompting of additional theories and speculative examinations of the data, are so rewarding that it appears worth adopting this difficult though less clearcut approach. It should of course be emphasised that the criteria of meaningfulness and probability levels will be applied repeatedly as tests of the causal models developed in this study, so that rigour will not be disregarded in assessing the acceptability of these models.

These past sub-sections have discussed the variety of statistical options which have been considered for this research, with the conclusion that regression techniques appear to offer the best methods of handling this particular data set.

Other methodological issues which are closely bound up with the analytical procedures used here are discussed in the following sub-sections. These are reliability, validity, significance, and age and time concepts. Sections 5.20 and 5.30 deal finally with regression and path analysis, which are the principal techniques chosen for data analysis in this study.

#### 5.14 Reliability and disattenuation

The reliability of the tests and other instruments used to derive variables for the analyses was seen to be of major importance, not merely because any research programme seeks to assess the reliability of the measures on which its conclusions are based, but also because some of the key analyses were based in part on the reliability data to correct for attenuation in the correlations prior to further handling of these relationships.

Ferguson (1971) points to the presence of both systematic and random error and shows that reliability estimates are based on the latter. It may well have occurred that E's assessments in one or more of the tests or interviews were systematically biased, since it was not feasible to control against a number of other E's. The fact that a single E gathered all the data means, however, that the correlation and regression analyses which form the basis of the study would not be affected by any systematic error which might exist. There are no conclusions based, for example, on a specific IQ assessment of the sample, or on other comparative data. The lack of a range of inter-observer reliability studies on each of the variables is thus not seen as a weakness of this research.

However, considerable emphasis needs to be placed on the concept of random error and the methods used in its estimation. It was decided to give much attention to this area. The fact that E administered virtually all the instruments, on all the children, was seen as an advantage in that the reliability determinations offered an estimate of combined instrument and observer reliability.

There are no conditions under which perfect reliability can be expected, where human behaviour is being measured. Even with perfect observers and instruments, the subject being tested has varying degrees of motivation for performance, so that there is no sense in which one could derive a 'true performance' from any one subject. As Cronbach et al (1972) show, there are at least three kinds of errors involved in making an estimate of a single person's universe score: error D, the difference between observed and universe scores; error d, the difference between the person's observed deviation from the sample mean and his or her universe deviation from the sample mean (which has a different variance from that of D); and error e, the ordinary regression error in predictions made on the basis of the person's observed score. The varying conditions of testing and the purpose of the tests mean that there is no sense in which one can refer to the reliability of a test. It is necessary to decide what is to be seen as error, and what is wanted information. Despite the conceptual difficulties, the authors warn of the importance of taking



reliability into consideration in studies based on correlations, by way of correcting for attenuation.

There are various factors within the instrument itself which also influence reliability coefficients. Brown (1976) describes these. They include the range of scores - as the variability of the scores decreases, the reliability decreases; thus a very small range tends to give a low reliability. The length of the test is related to this, and increasing the number of items increases the reliability. Test difficulty is another factor; here there is no simple relationship, but with tests that are too difficult or too easy for a group, the range of scores will be reduced and thus contribute to lower reliability.

It is thus clear that there is no simple approach to the problem of reliability. A test cannot be judged on the basis of the size of a simple  $r_{xx}$  figure without looking at these various considerations.

The interpretation and use made of reliability coefficients is not free of controversy. Bohrnstedt (1969) presents the usual formula for correcting a correlation for the unreliability of two variables, namely

$$r_{xy}^* = \frac{r_{xy}}{(r_{xx} \cdot r_{yy})^{1/2}},$$

but goes on to show that the correction of a raw regression coefficient involves only the unreliability of the predictor X, that is

$$b_{yx}^* = \frac{b_{yx}}{1 + \sigma_e^2 / \sigma_x^2}.$$

This is correct. A more basic formula for the regression coefficient supports Bohrnstedt's point:

$$b_{yx} = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{(N - 1) s_x^2}$$

Error in observing X or Y would not be likely to change the numerator radically. (Take the case where X is a categorical and thus reliable observation, and only Y observations are subject to error; the summation of  $(Y - \bar{Y})$  for each value of X would be the same no matter how serious the Y errors, provided  $\bar{Y}$  remains the same.) Clearly an increase in the variance of X,  $s_x^2$ , due to error in observations, is the only factor which would change the regression coefficient. But this argument becomes more dubious in a multiple regression situation. For example, with two independent variables we get:

$$b_{yx_1} = \frac{(\sum x_2^2)(\sum x_1 y) - (\sum x_1 x_2)(\sum x_2 y)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1 x_2)^2}$$

Again the squared error of  $x_i$  (squared deviation from the mean) affects the coefficient. But with an error-biased  $y$  now involved in multiple relationships the tenability of Bohrnstedt's position becomes more doubtful.

In other ways it is also an over-simplification. Many of the parameters derived in multiple regression situations rely on  $y^2$  and thus error variance in the criterion variable does influence the conclusions. For example, the multiple correlation, despite its conceptual limitations, relies absolutely on the variance of  $y$  in the determination of  $R^2$ . Thus,

$$R^2 = \frac{(\sum yy')^2}{(\sum y^2 \sum y'^2)^{\frac{1}{2}}}$$

shows the dependence on  $y^2$ . The determination of F ratios and regression sums of squares likewise involve the variance of  $Y$ . No complete analysis of a multiple regression situation is possible without considering error in observing  $Y$ .

This point needs emphasis in the light of Bohrnstedt's further argument that, in view of the different methods of estimating reliability, one should be cautious in interpreting coefficients that have been corrected for errors in measurement; he questions whether a correlation or regression should ever be corrected for unreliability. This applies particularly if one is aiming to predict a criterion, in the real world with a particular set of instruments. In contrast he recognises that a causal model does demand a situation of perfect (i.e. corrected) reliabilities.

The implication that studies for predicting a criterion should not attempt to take account of error, either in the predictors (the fallible independent variables) or in the criterion, is a way of ensuring the worst of both worlds; there is no way of judging the relative weights of predictors if their relationships with one or more criteria are left to chance error distortion, and if the fallible criteria (in different studies) are also left uncorrected.

The question of the method of determining reliability does demand close scrutiny in assessing or using a reliability coefficient. Two observers testing the same child at the same time, for example, might provide a correlation of observer reliability but would not offer any adjustment for instrumental unreliability. Such difficulties emphasise the need to examine and specify one's methodological framework in correcting for error, rather than to ignore the problem and offer predictions based on uncorrected correlations. As Cronbach (1976) points out, in a complex model aimed at ultimate prediction, neglect of the unreliability of correlations may well lead to totally incorrect predictions. Such an uncorrected model presumes a non-existent reliability.

This issue is discussed in more detail in a later section, 5.142, on

disattenuation of the working matrix.

One question that needs to be considered is how small a reliability can be accepted in deciding whether to use a test. Wiseman (1956) said that the concept of reliability was being inflated, perhaps at the expense of validity. Reliability was important but if its importance was over-emphasised one ran the risk of jettisoning useful and valid tests. The uncritical acceptance of lower limits such as .90, as a result of increasing expertise in devising objective tests of intelligence and attainment, could lead one to forget that tests with reliabilities of .5 or even lower may add significantly to a particular test battery. A highly reliable objective test may not always be highly valid.

This thinking is basic to the present author's approach to the analysis of the data. Clearly a test with low reliability is more open to question, but if it is thought or has been shown to have validity in terms of the other variables being studied, then the weight of judgement will be in favour of including the variable unless it is at an unacceptably low level. Naturally the weight that can be attached to a prediction from a variable with a reliability of only .4 to .6 will be much lower than that derived from one with a reliability in the range of .8 to .9.

The considerations voiced by Brown (*ibid*) emphasise how likely it is that certain kinds of tests, such as those scoring a maximum of only a few points, with most children scoring at one or other extreme, cannot but have a relatively low reliability. Yet the characteristic being scored may well help distinguish between performances on some other variable.

The choice of reliability method was virtually dictated by the nature of the research, with one person carrying out all the tests and nearly all the assessments. Ferguson (1971) describes the available methods - test-retest, parallel-form, split-half, and internal-consistency. Of these, the parallel-form requires the use of equivalent forms of the same test; these are not easily available. It was decided that test-retest offered the best means of assessing the majority of variables in the project. In cases where this would not be possible, the split-half or internal-consistency methods would be considered, or as a last resort inter-observer reliability. The Spearman-Brown is the indicated formula for split-half determinations, and Kuder-Richardson 20 or 21 for assessing the internal consistency of dichotomously scored items.

Bartko (1966 and 1976) and Bartko and Carpenter (1976) have offered serious criticisms of the use of the Kuder-Richardson and other well known methods of determining the reliability of intraclass correlations. Although their arguments and the intraclass measures are based on analysis of variance

methods rather than simple correlations between two sets of items, the authors are in essence questioning the whole concept of interpreting reliability as high in the case where one rater gives a group of five clinical subjects the scores 1,2,3,4,5, compared with another rater who scores the same subjects 5,6,7,8,9; equally the authors reject the calculation of high reliability where one rater scores 1,2,3,4,5, and the other 2,4,6,8,10 for the same subjects. Without entering here into the debate on the different statistical approaches of the workers cited by Bartko, it can be argued that in this case too one is faced with a situation rather like that defined by Wiseman (ibid), in which emphasis is placed on reliability rather than considering validity. Two raters who give their subjects exactly the same ranking can be seen as having a high degree of reliability in statistical terms. Their differing levels of judgement suggest, however, a serious lack of validity in the judgements themselves. One cannot somehow incorporate a validity factor into an issue of statistical reliability unless one is prepared to develop a new concept which estimates or combines both factors efficiently.

Questions of reliability will thus be dealt with here in terms of formal statistical agreement or correlation. Validity is treated as a separate issue, though one whose foundations rest on the framework of reliability.

Findings on the reliabilities of the variables are dealt with in section 6.50.

#### 5.141 Reliability and validity of interviews

Issues concerning test reliability are relatively straightforward and well charted compared to the problems of interview reliability. Many important studies based on interview data do not deal with this matter in any depth. At most there are a few statistics on inter-interviewer reliability. It is seldom that the accuracy of reported phenomena is assessed, let alone questioned. Even major text-books on interviewing techniques give at most a few pages to discussing the assessment of reliability, although a lot of attention is paid to methods of questioning which will minimise the degree of respondent inaccuracy and ensure a high level of accuracy in interviewer recording of responses.

This neglect is partly because of the confusion over the term reliability. The term 'validity' is often used to describe the closeness of the correspondence between the observer or interviewer's report and the responses voiced or intended by the respondent. This however is still essentially a matter of reliability. Validity can only be used to describe the extent to which the respondent (or interviewer's report) reflects the actual behaviours, attitudes or factual informa-

tion concerning the respondent or her children, for example.

It is possible to regard validity in the sense in which it is applied to test measurements as being very different from the above concept of interview validity. Test validity usually implies that the performances monitored in a test reflect to some degree an hypothesised underlying dimension; the Bender Gestalt measure of a child's ability to copy certain drawings is often interpreted as a test of visuo-motor integration, or at a deeper level as a means of assessing emotional or neurological problems. An interview may attempt to tap underlying attitudes and in that sense the term validity has a similar conceptual meaning. But when an interview is designed to assess past behaviours or facts known only to the respondent, the factual validity has a statistical form akin to reliability and is theoretically just as capable of being measured accurately, if other evidence is available on those past phenomena.

A thoughtful and early examination of the problems of interview measurement is presented by Maccoby and Maccoby (1954); they point to the contrast between the higher reliability of standardised interviews and the greater validity of the unstructured interview. Reliability in the ordinary meaning of the word is examined by Cannell and Kahn (1968) who quote a study showing significant differences between responses recorded by different interviewers, these differences often occurring in the direction of the interviewers' own opinions.

In the related area of the validity of respondents' reports on behaviours and background facts, both Cannell and Kahn (ibid) and Blum and Naylor (1968) present studies in which estimates of the accuracy of the reported facts are cited. Wenar (1963) summarises a considerable amount of evidence on the 'reliability' of children's developmental histories, as reported by mothers, starting with perinatal factors. Yarrow et al (1970) find very low agreement between the facts recorded in clinic and hospital reports and the maternal recall of those facts some while later.

Some of the above findings on the factual validity of interviews will be combined with other evidence on the reliability of the interviews themselves, in the section dealing with variable reliability (6.51).

As suggested in an earlier paragraph, factual validity in the sense of the accuracy of respondent recall or reporting on past (or current) behaviours is closely allied to reliability. The use of disattenuation methods has already been discussed. It seemed appropriate in this study to incorporate both the reliability of interview measures (so far as this could be ascertained) and the factual validity of this type of behavioural measure, as reported in the above studies or obtained from other sources, within a single reliability statistic for use in the disattenuation process prior to the regression analyses. Failure to take account of the fact that parental interview reports are only a partially

valid (or partially accurate) reflection of their educational behaviours towards their children would have the effect of reducing the importance of the parental variables in comparison with the disattenuated test variables. The alternative would have been to take no statistical account of either test reliability or the factual validity of interviews, a decision which would conflict with the strong warnings of Cronbach et al (1972) on the misleading results of analyses not corrected for the varying degrees of attenuation in the predictors.

#### 5.142 Disattenuation of the working matrix

The question of whether or not to disattenuate a matrix prior to deriving regression coefficients has been discussed on occasion in the literature, almost invariably by those favouring this procedure. With the exception of Bohrnstedt's criticism (described in section 5.14), which was based on the regression of a criterion on to a single predictor, there have been no substantial critics of this practice. The reasons why matrices are not generally disattenuated prior to regression or path analyses are somewhat obscure, since many studies present both reliability figures and regression analyses without explaining why no corrections have been made to the matrices on which the regressions are based. One possible reason is that well-known computer package programmes such as SPSS do not provide easy (let alone automatic) procedures for disattenuation prior to regression; another reason may be an understandable conceptual hesitance about increasing a set of correlations on the basis of error in the measurements of the contributory variables, since an intuitive but incorrect response to the finding of error would be to decrease correlations based on the error-laden measurements. The difficulty of obtaining reliability coefficients for all the variables in a complex model may be a further obstacle.

The arguments in favour of disattenuation are powerful and fairly consistent.

Cronbach (1976) has long urged the need for disattenuation. He points out that generalisability theory broadens the usefulness and interpretations of disattenuation. He cites Lord's evidence that correcting for error can at times change an ordinary partial correlation from positive to negative; he also cites Wiley's evidence of how path coefficients are altered considerably when they are corrected. An examination made by Cronbach and his colleagues of the important study by Bloom (1964) revealed that Bloom's finding that change in IQ from age to age had a slightly negative correlation with initial IQ (-0.05) was transformed into a sizable positive correlation when error in measurement was taken into account.

Cohen and Cohen (1975) deal with the same issue at some length. They cite the conclusions of Campbell and Erlebacher (1970) that failure to correct

for the fallibility of partialled variables in the analyses of the Headstart results may have led to mistaken conclusions as to the quality of the programmes. While the reliability of the criterion variable has some importance, it is the reliabilities of the partialled predictors which are crucial.

Cohen and Cohen state that the importance of correcting for error is so great that when no empirically based coefficients are available "We believe it is preferable to use informed 'guesstimates', particularly when they are low, (rather) than to make no correction for unreliability, which we have seen is equivalent to assuming  $r_{ii} = 1.00$ ". In situations of uncertainty it may be desirable to analyse the data on the basis of low and high estimates of the reliability of the unknown variables and then draw conclusions consistent with both sets of results.

McKennell (1977) points out that computer procedures for multivariate analyses typically examine a whole network of relations simultaneously; distortion of the pattern of the observed relationships due to measurement error thus becomes a serious possibility. However he suggests that while particular relationships should be corrected for attenuation, wholesale correction of an input correlation matrix is not to be recommended as a routine procedure without analysis of the underlying assumptions. This latter view does not appear to be shared by the authors quoted earlier. The arbitrary selection and correction of particular correlations in a matrix (implying increased values for those coefficients), while leaving the majority of the other interrelations unchanged, is likely to produce even more distortion than would the simple omission of any disattenuation.

Another worker who has stressed the importance of correcting for error prior to regression analysis is Duncan (1975). He warns that almost the whole of psychometrics can be seen as a frontal assault on the problems of error in variables.

The conclusion that can be drawn from the views of these different authors is that it is a legitimate strategy, based on sound statistical principles, to disattenuate all the 'working matrices' on which regression analyses are to be carried out in the present research. The alternatives of aiming for an almost error-free set of relationships under highly controlled laboratory conditions, or of relying solely on probability and significance statements to take account of measurement error, are unacceptable when more practical methods exist for coping with unreliability in field research.

One further advantage of disattenuation is that the procedure makes it possible to focus on the relative contribution of less easily measured variables which, despite the unreliability of their assessment, may be of equal or greater importance than those variables which can be measured more precisely; when such corrections are not made it is the more precisely measured variables which tend

to explain an undue proportion of the variance in a prediction equation.

Only a few different methods have been suggested for disattenuation. Bock and Petersen (1975), for example, put forward an algorithm based on introducing corrections for error down the diagonal and then examining eigenvalues of the resultant matrix; values below unity are altered to ensure that the matrix will invert successfully and a new matrix is then reassembled from the eigenvectors and the altered eigenvalues. The Bock and Petersen method is a rather complicated way of achieving what can be done more simply; moreover, increasing the low eigenvalues is a somewhat ad hoc approach which has no clear relationship with the reliability coefficients concerned.

Fuller and Hidioglou (1978) suggest an algorithm in which the derived regression coefficients are altered to take account of measurement error. Their method is interesting but in the context of the present research it would have been difficult to incorporate the V-ridge regression method (described in section 5.22) within their algorithm.

The basic method put forward by Spearman (1904) is still generally followed in disattenuation. The correction is given by

$$r_{x_i x_j}^* = r_{x_i x_j} / (r_{ii} \cdot r_{jj})^{\frac{1}{2}}$$

where  $r_{x_i x_j}^*$  is the corrected correlation

$x_i$  and  $x_j$  refer to the set of variables (predictors and criterion)

$r_{ii}$  and  $r_{jj}$  are the reliability coefficients for the variables  $x_i$  and  $x_j$

This correction has the effect of disattenuating all the correlations in the matrix, including those between the dependent and independent variables. For obvious reasons the correlations of unity down the diagonal (or the equivalent variances in the case of covariance matrices) are kept at their original values; correction would imply that a correlation greater than unity was possible.

There are a few serious problems which arise with the Spearman correction for unreliability. Goldstein (1979) shows that the use of reliability coefficients based on samples smaller than the sample whose matrices are disattenuated is a biased correction. This is a difficulty which cannot be overcome within the present study. Although variables with a wide range would enable split half reliability determinations on the basis of the full sample's scores, this would not be possible with the many short range variables.

Another problem is that any particular sample correlation coefficient has its own standard error; thus the corrections for measurement error within each of the variable constituents of the coefficient are applied to a value



which may well be at some distance from the mean value of a series of measurements of that coefficient. This problem is related to a third one, namely that in view of the potential error in the value of each coefficient prior to its disattenuation, it can happen on occasion that some corrected coefficients become greater than unity - an obvious impossibility.

The algorithm of Bock and Petersen (ibid) was intended to take account of this and other possible reasons for a failure of matrix inversion. But their method, as described earlier, implies alteration of the entire matrix on an ad hoc basis.

A simpler method, also ad hoc but affecting at most only a few coefficient values, was developed for the present study. This method has been built into the V-ridge regression programme described in section 5.22 and is applied automatically. A maximum value is set for disattenuated coefficients within the matrix, so that if any corrected coefficients exceed this value they are reduced to the chosen maximum.

Given the type of educational and psychological data collected for the project, it was decided to fix this maximum at 0.95. Such a correlation reflects a shared variance of 0.9025 between two variables and is probably around the maximum that is likely to occur within the framework of these variables - unless highly similar variables have been inadvertently employed. Previous research by the author supports the proposed maximum.

One further question concerns the reliabilities of the latent variables derived in the path analysis models. These latent variables were used as group variables representing the contributory predictors within each conceptual framework, such as the sets of cognitive skill variables and the sets of attainment and parent environment variables. Since the latent variables are each made up from a set of fallible predictors it was apparent that the latent variables themselves could not be seen as having a reliability of unity.

Accordingly the reliability of every latent variable was determined from the reliabilities of each of its constituent variables, using the following derivation:

$$r_{LV.LV} = \frac{\sum_{i=1}^{i=m} r_{ii} \cdot B_i}{\sum_{i=1}^{i=m} B_i}$$

where  $r_{LV.LV}$  is the estimated reliability of the latent variable  
 $r_{ii}$  is the reliability of each predictor variable in the set  
 $m$  is the number of predictors  
and  $B_i$  is the standardised regression coefficient for each predictor variable

### 5.15 Validity assessment

The concept of validity is a tenuous one, and yet it is of overwhelming importance in assessing both the generalisability of the research findings and the acceptability of the variables defined in that research. There are no firm rules for the determination of validity, although there is some agreement on the areas which should be considered. Whatever the nature of the research and the hypothesised variables, the assessment requires both concurrent and long-term study. The long-term study almost certainly involves replication of the research itself, under dissimilar conditions, and an examination of the hypothetical variables over widely differing areas.

The concurrent analysis of validity requires a consideration of the degree to which any conclusions from this research, positive or negative, are generalisable to other schools, children and parents; it also requires a review of the meaning and validity of the individual variables, how they behave in a situation of complex interrelationships and how they contribute to the criteria of reading and mathematical performance.

Campbell and Stanley (1963) have written one of the best known studies of experimental and other research designs. They list 12 criteria by which the generalisability of a design can be judged.

Of the eight factors affecting internal validity, the design of the present study will not be affected by differential history, maturation, test-retest factors, or changes in instrument or observer, since all children in the sample experience these equally; only the intervention programmes themselves differentiate the group histories. The question of whether there is differential attrition is dealt with in section 6.10. Selection-maturation differences are related to the questions of differential selection bias and statistical regression to the mean; these three issues need to be considered carefully in the light of the problems experienced in obtaining suitable control groups. The randomness or otherwise of the sample groups is discussed more fully under sections 4.222 (distribution of nursery children), 4.224 (selection of schools) and 4.226 (selection of parents), while the size and nature of the differences between the research groups is analysed in section 6.20.

Looking next at Campbell and Stanley's four factors affecting external validity, it appears unlikely that there would be a nursery pre-test interaction with the intervention programme. Nor are there multiple treatment or experimental arrangement effects (apart from the programmes themselves) which could affect the groups differentially. However, the possibility of interaction effects between the experimental programmes and the research groups merits

serious consideration, and is linked to the discussions noted at the end of the previous paragraph.

Using the Campbell and Stanley nomenclature, the present research can be described as a quasi-experimental design with non-equivalent control groups, since it was not possible for children to be allocated randomly into groups.

The validity of the hypothesised variables needs to be examined in a different way. These variables are conceived on the basis of a variety of tests, measurements and interview questions. Various authors (King, 1969, Kerlinger, 1973, Pidgeon and Yates, 1968) present the main headings under which validity is usually assessed: content or face validity; concurrent validity (comparing the measurement with other forms of assessment at the same time); congruent validity (comparisons with other tests known to have high validity); predictive validity; and construct validity (often analysed in terms of loadings on factors thought to represent certain psychological constructs).

Because the concept of validity is so tenuous there are many different interpretations. Lindeman (1967) suggests that the old adage, "validity is determined by whether a test measures what it purports to measure", should be replaced by the concept of asking what does a test measure, and then deciding whether a test meets a particular need. In reply to criticism of construct validity, because of its frequent reliance on 'factors', Campbell (1960) points out that many tests are validated only by predictive methods; yet it is rarely that tests can serve as (accurate) predictors of later performance in a setting where there are complex determinants of success.

The question of personality assessment is one of the most difficult in this area of validity. As Cronbach (1961) points out, there is much debate over whether performance tests are acceptable for the measurement of personality. While the highly structured nature of the test situation does ensure uniformity of conditions and even similar motivation to some extent, this structured situation poses questions of generality in other non-test situations. Cronbach also shows that the development of a single test will not necessarily isolate important individual differences. He concludes that highly structured tests of narrowly defined variables have little usefulness except in the development of psychological theory; even then they are not of much use except as part of a composite or battery. Since numerous traits interact to determine behaviour in any situation, one such measure will rarely have a large correlation with any practical criterion.

Cronbach's thinking is basic to the approach towards all measurement in the present research. Many of the tests used here are attainment or ability tests in which personality variables are thought to play a moderate part. A

few of the measurements aim at a partial assessment of traits such as reflectivity, self-esteem or distractibility. However, none of the tests is seen as representing any narrowly defined characteristic. Rather they have been included because, despite their coarse-grained nature, they have been shown in previous research (or are currently hypothesised) to serve as contributory predictors to the criteria of interest, namely reading and mathematics, and also because they are thought to be conceptually related, in one or other way, to these same criteria. In so far as the test behaviours appear to represent the hypothesised abilities or traits, they are used and defined provisionally as such, although it is clear that any definition here can only be partial.

Elsewhere, Cronbach and Meehl (1955) discuss general concepts of validity and offer a radically new idea which has drawn little practical support in the intervening years. They argue in favour of a much wider interpretation of construct validity, although they warn that it is rarely possible to derive a numerical coefficient for this interpretation - perhaps only a set of upper and lower bounds can be quantified. What they offer is the idea of a nomological net - an interlocking system of laws constituting a theory in support of the construct validity. In other words, the construct has to occur in a nomological net, at least some of whose laws are observables and which shows explicit steps of inference. Thus rationalisation cannot be seen as construct validation; it is essential to have a rigorous chain of inference and clear hypotheses as to the induction from data.

Cronbach and Meehl recognise that because of the essential vagueness of psychological laws it is necessary to work with crude, half explicit formulations and vague incomplete networks lacking the rigour of calculus, for example. And yet it is these networks which give the constructs whatever meaning they do have. One does not experiment to 'prove the theory and validate the test', but rather one predicts a relationship between test scores and other variables, gathers data and if the prediction results in harmony, one can retain the belief that a test measures the construct. It can never be proved. Many types of evidence are relevant to construct validity, including content validity and inter-test correlations.

Brown (1976) is among the few authors to have developed the concept of the nomological net. He also builds on Campbell's model of a multitrait-multimethod matrix.

This brief review of the field provides a necessary background to the arguments which follow.

In most cases test validity is defined in terms of single correlations between the test itself and a variety of other measurements, both within the

same conceptual area and across boundaries into related areas (for example, Brimer and Dunn, 1962; Wechsler, 1967; Koppitz, 1975). These methods have a certain adequacy in that they justify the claims of an individual test to be used as a method of assessing a particular behaviour or ability. But they present a one-sided picture of their predictive power, because almost any criterion is determined by a large number of variables, such as skills and traits, environmental conditions, etc. A claim for a test correlation of .70 with reading performance a year later simply means that on its own the test in question shares half the variance of the criterion variable. Several variables from other conceptual fields could in combination offer a multiple correlation of .80 with year-end reading, and the inclusion of the previously meaningful test might add no more than one or two points to that prediction.

There is also the problem of spurious correlation. Some antecedent variables - cognitive abilities and high motivation, for example - may combine to create high performance on tests of both reading and skilled problem-solving. The correlation between problem-solving and reading would offer no evidence of causation; even the argument that problem-solving skill was a good predictor of reading would blur the reality of their joint indebtedness to other factors.

There are moreover many factors which are not usually assessed in determining test performance. The uniqueness of the test situation and testers, the child's memory, motivation and state of stress, and the interaction between child and tester all influence performance on tests - often to a varying degree depending on the nature of the test.

In what could be an area either of brash over-simplification or insoluble complexity, it is necessary to arrive at some clear and practical principle on which this study will approach the question of measurement validity. This will be done along the lines suggested by Cronbach and Meehl, for examining validity within a nomological net. There will be three specific forms of this validation:

- i. Content validation, based simply on the description of the measurement procedures and the rationale for their inclusion in a test battery or interview protocol.
- ii. The contribution of the variable measurements to the total analysis of the interlinking areas in the intervention programme.
- iii. The assessment of each variable in terms of what might be newly defined as its nomological validity.

This last form of assessment is in effect a highly specific method of measurement within the nomological net. It presents the collected details of

four nomological parameters of each variable. It needs to be emphasised that such a statistic is an assessment of a variable within a particular hypothesised model. Unlike the ordinary validity assessments which look only at simple relationships between the variable and a criterion or parallel construct, this assessment attempts to portray the variable's relative importance within the model.

The four parameters are the following:

- a. The single contribution of variable  $i$  to the criterion, as measured by the simple correlation squared; this is equivalent to the contribution made if that variable is entered first into the regression equation for the criterion. It can be defined as  $r_i^2$ .
- b. The total contribution of all the variables used in the model. This is normally defined as  $R_t^2$ , the squared multiple correlation of all variables with the criterion. This is a key indicator in validity data because it shows how completely or incompletely the model succeeds in locating the theoretical contributors to the criterion. For example, high additive validity for a variable (as in (d) below) may be less impressive if only 20 per cent of the criterion variance has been accounted for in the model.
- c. The unique contribution made by variable  $i$  to the multiple regression equation, when it is entered into the equation alongside all the other variables in the model. This can be defined as  $R_t^2 - R_{t-i}^2$ , with  $R_{t-i}$  being the multiple correlation of all the variables (with the criterion) except variable  $i$ .
- d. The additive validity, which is defined here as the ratio of the unique contribution to the single contribution, namely

$$\frac{R_t^2 - R_{t-i}^2}{r_i^2} .$$

This last ratio is a key statistic because it compares the ultimate importance of the variable, when all the other variables are also put into the model, with the apparent importance when the variable is judged as the only contributor to the criterion. The latter relationship is of course only a laboratory definition, for it is seldom that any criterion is dependent on a single predictor in the social sciences.

The nomological validity of a variable can therefore provide a set of four parameters to interpret in detail the validity of the variable within a particular model. To these can be added two further parameters of interest, namely the regression coefficient and the probability of this coefficient. Normally the validities of all the variables within a model would be presented at the same time, as this also helps to define the validity of any particular variable. Two typical examples are presented in table 2 overleaf.

Table 2. Presentation of nomological validity within two typical models

<b>Model MN2. Prediction of post-test numeracy from pre-test scores</b> N = 129      Total variance predicted .34      Shared variance .12					
Variable	Single contrib.	Unique contrib.	Additive validity	Standard. regr.coef.	Probty. of coef.
Maths numeracy Nursery	.45	.128	.28	.22	.000
Maths concepts Nursery	.24	.040	.17	.13	.000
Piag. tests Nursery	.11	.006	.06	.06	.01
Reading aware. Nursery	.12	.005	.04	.05	.005
Infant Read. Test Nur.	.25	.041	.16	.13	.000
<b>Model R1. Prediction of post-test reading from pre-test scores</b> N = 48      Total variance predicted .28      Shared variance .10					
Variable	Single contrib.	Unique contrib.	Additive validity	Standard. regr.coef.	Probty. of coef.
WPPSI Inform. Nursery	.11	.031	.27	.15	.02
WPPSI Pic.Com. Nursery	.16	.014	.09	.09	.08
WPPSI Block D. Nursery	.12	.013	.11	.10	.08
Rhythmic tap. Nursery	.13	.034	.27	.16	.01
Bender Gestalt Nursery	.10	.004	.04	.05	.22
Self-picture Nursery	.10	.004	.04	.05	.19
Sex	.09	.032	.38	.18	.02
Age Nursery	.12	.045	.38	.19	.005

Both models are typical research situations in which sets of variables, in raw or standardised form, are used to predict the scores of criterion variables. A comparison of the models in terms of their nomological parameters enables various characteristics to be examined which are not normally apparent.

Within each model the additive validity figures (the ratio of the unique contribution to the single contribution) provide a comparison of the relative usefulness of each variable in the model. Decisions on the retention or rejection of a variable can then be made not only in terms of the usual assessment of unique contribution (addition to  $R_t^2$  when entered last) and significance of the regression coefficient, but also in relation to the nomological power of the variable in so far as its unique contribution to variance bears a reasonable relationship to its single contribution.

Despite the fact that the conventional indicators of total variance predicted and probability levels suggest that MN2 is a better model, the nomological validity indices show that R1 is in fact more efficient. An examination of the single contributions and the additive validities explain the difference. Model R1 has a number of predictors with relatively low single contributions (the highest correlation with the criterion being only about 0.40) and most of these predictors make an efficient unique contribution to predicted variance, despite the modest significance figures for the regression coefficients.

In contrast the variables in model MN2 display some relatively high single contributions alongside some rather modest additive validities. In other words, the latter model is beset by higher levels of multicollinearity, which have the effect of reducing the unique contributions of predictors. The highly significant probabilities for the model's regression coefficients do not overcome the fact that there is more unused variance or 'redundancy' in MN2 than in R1. This is no reason for rejecting or questioning MN2 as an acceptable research model, but it does focus on a new approach towards assessing the quality of models of relationships, in a wider context than the rather limiting criteria of probability and total variance predicted.

In the past much regression analysis has tended to stress the size of the variance predicted as an indication of the credibility of the hypothesised model. Goldstein (1976) has warned of the serious limitations of such an approach. Even today computer package programmes continue to highlight the parameter of total variance predicted and the related F-statistics on the significance of additions to that variance. The result is that analyses presented in the literature often fail to tackle or speculate on deeper issues in the network of relationships and seldom examine the efficiency or frugality of an hypothesised model.

The validation tables presented in section 6.52 will show various examples of both acceptable and unacceptable models, judged in terms of their nomological parameters.

The usefulness of the nomological concept suggests the need for further research into its potential. At this point, however, it is necessary to stress that as with any statistical criteria where repeatability of results is essential to dialogue and a comparison of findings, the regression algorithm used to derive the validity statistics will also need clear specification. The present research study has based all its regression analyses on V-ridge, a method developed to minimise the unreliability of ordinary least squares on small and moderate sized samples (section 5.20 et seq.)

Of the three forms of validation specified in a, b and c, earlier in this section, content validation, (a), has already been defined in the description of



the tests and interview measures (sections 4.31, 4.32 and 4.34, and Appendices A and B).

The contribution of variables to the total analysis, (b), will be covered in sections 6.70 and 6.80.

The nomological validity, (c), and a few findings about individual tests and measures will be presented in section 6.52 et seq.

#### 5.151 Rationale for a Redundancy Index

Following the development of indices of the nomological validity of variables within a model - the key innovation being the concept of additive validity - it was considered desirable to develop a further statistic which would reflect the overall performance validity of a regression model, taking into account features of the model as a whole as well as the behaviour of its constituent variables. What was sought was a combined statistic which would reflect quantitatively what might be evident to the practised eye, namely whether there was a degree of superfluity in the predictive power of the individual variables, whether and to what degree each of the predictor variables made its own useful contribution to the model, whether there was a high degree of multicollinearity and whether the model offered a reasonable level of prediction of the particular outcome variable. These can all be seen as aspects of the validity of the model as a whole and are closely related to Cronbach's concept, previously referred to, of nomological validity.

For a variety of reasons, mainly statistical, it was thought preferable to express the combined validity measure in the form of a redundancy index. This index may be said to incorporate each of the statistical criteria which should normally be taken into account when deciding on whether to include or exclude variables in a regression equation. In essence the redundancy index provides a quantitative assessment of the superfluity and lack of economy in a multiple regression model.

In the early stages of developing this index a relatively simple interpretation of redundancy, based on the inverse of the mean value of the additive validities of variables retained in the model, proved inadequate to the task of defining the strengths or weaknesses of differing models. A number of alternative parameters were created in terms of various theoretical perspectives, then combined into single indices and tested. Many of these combinations were apparently successful but failed to yield a high statistic for those models which were clearly troublesome, or alternatively the statistic penalised some evidently satisfactory models. The statistic in its final form was tested on a variety of models and appears to offer intuitively acceptable values for these models.

The combined index is made up of four parameters:

i. Variable invalidity. This is the inverse of the mean additive validity of the individual variables used as predictors in the model. It is represented by

$$1 / \left[ \sum_{i=1}^{i=k} (u_i^2 / r_i^2) / k \right]$$

where  $u_i^2$  is the unique variance predicted by each variable within the model,

$r_i^2$  is the squared correlation between the variable  $i$  and the dependent variable, and

$k$  is the number of predictor variables

The nature of the combined redundancy index requires that the mean validity should be in the denominator of the above statistic, since an increased level of additive validity for each of the contributing variables implies a reduced level of redundancy in the model as a whole. Although this parameter has some links with the second parameter in the combined index, it plays its own part by providing an 'average' value for all the variables included in the prediction; variables with little or no independent contribution to the model will increase considerably the value of this parameter of mean invalidity. Under conditions of perfect validity for each of the predictors (in an orthogonal design, for example), the statistic will have its lowest value of unity.

ii. Unrealised predictive power. This is the ratio of the sum of the individual 'predictive' variances in a model to the total variance predicted by the variables acting in combination. It is defined as

$$\left( \sum_{i=1}^{i=k} r_i^2 \right) / R_t^2$$

where  $R_t^2$  is the total variance predicted by the model (corrected for the number of cases and number of variables)

This statistic on its own reflects the degree of 'wastage' in the model in relation to the number of variables used, and is linked to the problem of over-prediction or multi-collinearity. If the variables are orthogonal the summation of  $\sum_{i=1}^{i=k} r_i^2$  will equal  $R_t^2$  and there will be no loss of predictive power. Under such conditions this parameter too will have its lowest value of unity.

iii. Predictive shortfall. This is the inverse of the overall variance predicted in a model, and is given by

$$1 / R_t^2$$

It is a simple statistic, reflecting the overall predictive power, or lack of power, in a model. The inverse is employed to emphasise the redundancy in

those models having very little predictive power. Under the condition of perfect or total prediction this statistic will also have a value of unity.

iv. Inverse of number of predictors. This parameter, given by

$$1 / k$$

is a necessary corrective, for the reasons explained below.

The combined redundancy index is derived from the product of the four parameters defined above. To reduce the range of this statistic, the logarithm of the product is used as the index, giving:

$$\begin{aligned} \text{Redundancy Index} &= \ln \left[ \frac{1}{\sum (u_i^2 / r_i^2) / k} \cdot \frac{\sum r_i^2}{R_t^2} \cdot \frac{1}{R_t^2} \cdot \frac{1}{k} \right] \\ &= \ln \left[ \frac{\sum r_i^2}{\sum (u_i^2 / r_i^2) \cdot R_t^4} \right] \end{aligned}$$

It can be seen that an increase in the value of the index represents an exponential increase in the redundancy within the model, equivalent to a decline in its overall validity. With an ideal prediction of unity by a set of orthogonal predictors - a most unlikely situation in the social sciences, where there is nearly always some overlap of predictive power between variables and a less than perfect predictive model - the values of the first three constituents of the redundancy index would all be unity, giving a combined index value of  $(0 - \ln. k)$ , where  $k$  is the number of predictor variables.

The reasons for including  $k$  in this index need to be examined.

In a properly constructed model each variable added to the model should help to define the complexity of the contributors to that model, provided every variable makes its own meaningfully large (and statistically acceptable) contribution to the total variance explained. Unfortunately the extension of the model to include more 'explanatory' variables is usually offset by a rise in colinearity effects and a decline in the additive validity of variables already in the model.

This phenomenon is of course well recognised. What is not so generally accepted is that the addition of complexity is an acceptable loss if it also brings both added comprehension and added prediction to the model. The Occam's Razor principle (Edwards, 1972), previously discussed, is invoked by some as though simplicity was in itself desirable. In fact the Occam's Razor principle has its justification in choosing the simpler of two comparable models rather than in the rejection of the more complex of two competing or different models.

The adaptation of the redundancy index to take account of the 'complexity' consideration proved puzzling. When based only on the product of the first three parameters the index consistently favoured models with only two or three predictors, where there was clearly little competition to affect their contributions. Even the low level of overall prediction (the third constituent parameter) could not overcome this distortion in the performance of the index.

When the combined index was however corrected by dividing the product of the first three parts by the number of predictor variables in the model, the index appeared as a highly credible statistic with which regression models could be compared - taking into account each of the four characteristics built into that index.

### 5.152 Examples of the Redundancy Index

Five examples of the redundancy index statistic and its constituent parameters, taken mainly from the regression models used prior to the path analyses in this study, are presented below by way of illustrating the performance of the index.

Table 3. Prediction of post-test reading attainment (disad. reading groups)

Independent variables: Three pre-test attainment variables remaining after initial regression based on all five pre-test attainment scores				
N = 48				
Total variance predicted: 0.424      Shared variance: 0.144				
Predictors	Single contribution ( $r_i^2$ )	Unique contribution ( $u_i^2$ )	Additive validity	Standard. regr. coefft.
1. Reading awareness	.182	.076	.421	.200
2. Infant Reading Test	.416	.191	.458	.347
3. Maths Numeracy	.137	.013	.096	.084
	Variable invalidity	Unrealised pred. power	Predictive shortfall	Redundancy Index
Redun. Index parameters	3.08	1.73	2.36	1.434

This is clearly a satisfactory model with a modest redundancy index. The prior removal of two poor predictors helped create an adequate equation with two highly 'valid' contributors and one moderate contributor. The comparison with a similar predictive model (overleaf) shows some importance differences.

Table 4. Prediction of post-test total attainment (disad. Black children)

Independent variables: All five pre-test attainment scores				N = 44
Total variance predicted: 0.412		Shared variance: 0.127		
Predictors	Single contribution ( $r_i^2$ )	Unique contribution ( $u_i^2$ )	Additive validity	Standard. regr. coefft.
1. Reading awareness	.253	.089	.350	.212
2. Infant Reading Test	.203	.049	.242	.171
3. Maths Numeracy	.407	.143	.351	.269
4. Maths Concepts	.107	.002	.016	.031
5. Piagetian Tests	.034	.002	.074	.039
	Variable invalidity	Unrealised pred. power	Predictive shortfall	Redundancy Index
Redun. Index parameters	4.83	2.44	2.43	1.740

This redundancy index is (prior to the logarithmic transformation) approximately 36 per cent higher than that of the previous model. It is apparent that the presence of two variables with very low unique predictions (.002 each) is the main flaw in this model. Since the equation is one of the 'satellite models' based on a wider sample (satellite models being used to enable comparisons between sub-groups, with certain variables being retained in both sets of equations

Table 5. Prediction of post-test total attainment (disadvantaged girls)

Independent variables: Either of the pre-test ability variables remaining after removal of sex variables and two poor predictors in the ability group				N = 71
Total variance predicted: 0.496		Shared variance: 0.175		
Predictors	Single contribution ( $r_i^2$ )	Unique contribution ( $u_i^2$ )	Additive validity	Standard. regr. coefft.
1. WPPSI Information	.261	.034	.129	.142
2. WPPSI Sentences	.185	.046	.250	.175
3. WPPSI Picture Compl.	.249	.078	.313	.211
4. WPPSI Block Design	.249	.056	.227	.200
5. Rhythmic tapping	.027	.012	.446	.092
6. Matching Fam. Figs.	.073	.016	.224	.108
7. Bender Gestalt	.075	.001	.009	-.021
8. Age Nursery Assess.	.151	.078	.515	.237
	Variable invalidity	Unrealised pred. power	Predictive shortfall	Redundancy Index
Redun. Index parameters	3.78	2.56	2.01	0.894

to facilitate the comparisons, as will be explained later) it was not possible to remove these poor predictors. Thus, not only is the variable invalidity parameter worse (higher), but so too is the unrealised predictive power in the equation. Both models however predict approximately equal proportions of the dependent variable.

Table 5 (previous page) is an exceptionally powerful model, with a very low redundancy index despite the presence of one poor predictor, Bender Gestalt, which has no real contribution in this model (its negative regression coefficient, contrasting with a positive correlation coefficient, is a type of statistical artefact which will be discussed elsewhere in this study). Again the poor predictor has been retained since this too is a 'satellite model', enabling a direct comparison between boys and girls. The first two redundancy parameters are not particularly low - variable invalidity and unrealised predictive power - but since the model contains eight explanatory variables, of which seven make credible contributions to the prediction, there is an intuitive acceptability about the equation and this is reflected in the satisfactorily low level of the redundancy index. Visual examination of the table shows how each variable (with the one exception) makes its meaningful contribution alongside the other competing variables.

Table 6. Prediction of post-test total attainment (disadvantaged boys)

Independent variables: The two Nursery Needs variables				N = 58
Total variance predicted: 0.076		Shared variance: 0.012		
Predictors	Single contribution ( $r_i^2$ )	Unique contribution ( $u_i^2$ )	Additive validity	Standard. regr. coefft.
1. Need for Security	.081	.035	.432	.158
2. Need for Esteem	.075	.030	.396	.145
	Variable invalidity	Unrealised pred. power	Predictive shortfall	Redundancy Index
Redun. Index parameters	2.42	2.04	13.07	3.474

The equation shown in table 6 is a most unsatisfactory model, with a redundancy index (prior to the logarithmic transformation) about 13 times higher than the index for the previous model. It should of course be noted that the two models are concerned with totally differing sets of predictors and outcomes, and would not normally be compared other than in this methodological context. It is clear what the reasons are for the poor performance of the latter model. Only two variables, each with limited predictive power, are employed. It is interesting to note that in this limited context they each have a reasonably high additive validity; as a result the variable invalidity parameter is quite low.

There is also little unrealised predictive power, as can be expected with so few variables. The serious deficiency in this model is in its predictive shortfall. Since the equation has in fact been used in one of the path models, despite its shortcomings, it should be explained that it does not appear as an independent model but serves merely as a preliminary stage in the creation of the wider path model.

Table 7. Prediction of post-test total attainment (disadvantaged White children)

Independent variables: The two latent variables remaining after removal of three poor predictors in the group of Nursery Status predictors N = 62				
Total variance predicted: 0.720      Shared variance: 0.100				
Predictors	Single contribution ( $r_i^2$ )	Unique contribution ( $u_i^2$ )	Additive validity	Standard. regr. coefft.
1. Nursery Ability	.598	.256	.428	.360
2. Nursery Attainment	.686	.365	.532	.430
	Variable invalidity	Unrealised pred. power	Predictive shortfall	Redundancy Index
Redun. Index parameters	2.08	1.78	1.39	0.949

Table 7 shows another exceptionally low redundancy index, and again it is evident why this should be so. Both latent variables fulfil their purpose by making meaningfully high contributions to the total prediction at this level in the path model. It is worth noting that the additive validities of these two variables are much the same as the additive validities for the two variables in the previous model, giving a variable invalidity parameter not much different from that of the earlier model. The unrealised predictive power, as a ratio of the combined single contributions (or squared correlations with the outcome variable) to the total variance predicted, is also much the same in each model. It is the great difference in the total predictions which is responsible for the equally major differences between the redundancy indices for the two models.

What needs to be emphasised at this point is that the redundancy index is a valuable statistical pointer to the power and credibility of a regression equation. But it should always remain as a guide when considering competing models, rather than as a final determinant of the acceptability of a model. Each of the five models described here has its own usefulness and legitimate place in the total examination of the data from this study, although under other conditions a high redundancy index as well as poor additive validity figures might suggest alteration of one or more of the equations. The indices should serve thus as warning indicators, in the way that probability figures should do. The lower

the additive validities and/or the higher the redundancy index, the more caution should be observed in adhering to a particular model. At the same time such cautionary indicators cannot offer direct answers as to the inclusion or exclusion of variables from a model, nor can they replace the need to build each model carefully in terms of its own conceptual demands and limitations.

The redundancy indices of the models used in the study are presented in section 6.522 and in section 6.80, alongside the path analyses.



## 5.16 The significance concept

Educational and psychological literature is dotted with references to 'significance levels' and the implications of using this criterion as a benchmark for assessing the acceptability of research findings requires therefore some consideration, particularly in view of the sharp attacks that have been made on the criterion by various authors.

A fundamental criticism of the use of this criterion is that it often serves to justify 'decisions' about the acceptability or otherwise of an hypothesis. Tukey (1970) criticises the attempts to use decision theory to embrace all statistics. He suggests rather the use of tentative 'conclusions' about what has been indicated by the data. Thus rather than attempt to 'decide' on the 'truth' or otherwise of an hypothesis, one should choose wisely among possible alternatives, without being too certain about ones judgements. The value of conclusions is that they enable the experimenter to take into account all the factors involved in the study rather than rely on a single statistical criterion.

A number of writers in the educational, psychological and sociological fields offer various criticisms of the misuse of the significance criterion.

Bakan (1970) points to the widespread misunderstanding of the meaning of significance. It is not the probability that the results are due to chance, nor is it evidence of the 'goodness' of the inference or of the confidence with which the results of the experiment can be viewed as repeatable. The p value within the inference model - that is, the value set by the experimenter - is simply the value of the improbability of an event under the null hypothesis. Another way of formulating this is that the significance level is the probability of finding a certain result, given that the null hypothesis is true.

Bakan considers that much mischief has been associated with the use of the test of significance. There is a great deal of evidence that, given sufficient numbers, almost any relationship can attain a very low p figure; because of the reliance placed on high significance figures it is probable that many Type I errors are incorporated in the body of 'accepted' research findings (Type I errors meaning claimed success when the results are in fact really due to chance). Bakan refers to the proposal for using a loose null hypothesis, with the criterion of material significance, that is, testing the hypothesis of a deviation by a stated amount from the null point.

Nunnally (1970) argues that estimation is a far more reliable indicator of the value of a model than hypothesis testing. If a variety of indicators such as mean differences, correlations, the form of the curve of relationships and

other parameters are used, one can reach practical conclusions about a set of findings. He sees analysis of variance as essentially an estimation device for the different effects and their interactions. We need to ask what proportion of the variance do they explain rather than how large are they. Rejection of the null hypothesis is only a minor though necessary question.

Further criticism is offered by Meehl (1970) who points to a variety of methodological problems in the 'decisions' taken in relation to tests of significance. Slater (1978) compares significance tests with the use of Bayesian criteria; the latter do not attempt to give us clear cut inferences nor do they requires us to accept or reject anything. Bayesian analysis simply updates the researcher's opinions in the light of new evidence.

Many of the criticisms of the use of significance criteria are examined in a volume edited by Morrison and Henkel (1970) in which both sides of the debate are presented. Arguments in favour of the use of the test of significance centre on its value as a widely accepted criterion for the acceptability or not of research findings; shortcomings in the use of the criterion cannot be blamed on the criterion but on the misunderstanding of its purpose and the failure to recognise that large  $N$  implies the need to lower the probability level at which the null hypothesis can be rejected. Winch and Campbell (1970), for example, consider that it is very important to have a formal and non-subjective way of deciding whether a given set of data shows haphazard or systematic variation; if the test of significance shows that the data indicate systematic variation the analysis is not concluded but is just getting under way; it should not however be left to the investigator to determine what is systematic or haphazard.

Morrison and Henkel themselves come to highly critical conclusions on the basis of the evidence presented by the different authors. Among specific points they cite the irrationality of the choice of particular levels of significance; such levels can only be set rationally when the cost of a wrong decision about the null hypothesis can be estimated. The failure to consider the power of a test (the power against a specific alternative hypothesis being that proportion of the  $H_A$  distribution falling within the  $H_N$  region) means that the choice of level is practically without meaning, since the rejection of  $H_N$  is increasingly guaranteed as samples become larger; in contrast the power criterion is conservative and takes into account the size of the sample. The many publication practices (some journals for example publish only results which achieve levels that are 'highly significant'), the habit of searching for and then reporting only the significant findings from a large body of findings, and the failure to consider the joint significance of a set of findings are also criticised.

The authors conclude that the decision making process involved in reliance

on tests of significance is antithetical to the information accumulation process of scientific inference. Generally no action follows on 'decisions' about the  $H_N$ ; moreover such decisions tend to discourage more appropriate interpretation of the data. They suggest that the tests have little place in modern psychology and point out that they are seldom used in related disciplines such as econometrics, except where the cost of a wrong 'decision' is quantifiable. They argue that the use of the tests should be limited to a set of conditions which they specify; these include situations where the costs of a wrong decision can be calculated so that the level of significance can be rationally set; where the power of the test can be determined; and where causal inference is either irrelevant for the research purpose or unequivocal within the research design.

In the light of these considerations all the findings in the present study will be assessed according to two main criteria. The first will be the levels of probability yielded by the tests or statistical procedures which produce those findings. The second will be the size of the differences or unique variances whose probability levels are being described. Significance levels as such will not normally be set, among other reasons because a very considerable number of parameters will need to be described.

At the same time it is recognised that p figures of .05, .01 and .001 will retain their conventional meaning in so far as judgements may need to be made about certain parameters. Within the path analyses, however, varying criteria of probability levels and minimum unique variances will be employed, related to the N within each individual model, to determine which variables should remain in the model. In every case the criterion will be spelt out formally.

### 5.17 Age and time concepts

Concepts of age and time were inextricably woven into the whole project and it was important to consider in what ways they could be taken into account in the analyses.

For each child there were four basic age variables: cohort age, that is, the child's age in relation to the whole sample; the age of administering the initial battery of cognitive measures; the age of the mid-test battery of cognitive measures; and the age of the post-test assessment of reading and mathematics. The principal time considerations were those of the time spent by the child in the nursery class after the initial assessment of cognitive skills, the time spent in the reception class between entry and the date of the post-tests on that child, and the length of time that the child's mother attended a parent programme. A more accurate measure of the last-named variable was available in the form of scores for meetings attended or partially attended.

The number of these temporal variables might have been reduced had it been possible to formalise the intervention as a laboratory type experiment in which the age span of the children was limited to a few months at most; it would also have been necessary to test all the children within the same brief period, say two weeks for any one battery, and to select only those children who would transfer from nursery to reception class at the end of the same term. This would have required a much larger number of sample schools to provide the same number of children and would have demanded the assistance of several other researchers to conduct the three test batteries within the same brief periods. Exclusion of most of the temporal variables (homogeneity in parameters eliminates the possibility of using their variances in analysis) would have had the further effect of obscuring their possible explanatory power.

As the statistical model planned for the main analyses was based on regression and path analysis, it was considered that a diversity of age and time variables could well be taken into account within a less tightly controlled field research situation.

There were however a number of problems which needed examination. As Goldstein and Fogelman (1974) point out, alongside the ordinary concept of standardising tests of school attainment according to the child's test age, there is the fact that the average test score for any particular mean age differs according to the time of the year. A separate paper (Goldstein, 1972) points out that the variance of a growth variable such as height itself varies with age, even over a relatively narrow range of ages.

The analytical design of the present project was able to take account of

the ordinary variation with test age. Although there was a range of several months in the time of the year during which each of the batteries was administered, this range was small enough to minimise the likelihood that seasonal or school term effects could have a radical influence on scores. The third problem raised by Goldstein, the change in the variance of a growth variable according to the age of measurement, was a more serious one. Figures cited by that author indicate a change of 3 per cent over 12 months in the size of the variance of height at the age of 7 years.

For younger children the alteration in variance over age could be considerably greater. The difficulty in taking account of this problem in the present research was that each of more than 50 variables might have its own parameters of change in variance with the age of the children concerned. It would have been extremely difficult to take account of this form of change, let alone to undertake the research necessary for its assessment. It was thus a problem of which cognisance was taken, but not one which could be mastered within the ordinary multivariate model. Since many of the changes in variance were likely to be in the same direction for conceptually similar variables, it was thought that distortion of the results might well be minimal within a path model.

A more fundamental problem concerning age variables has been raised by Gewirtz (1969). In his theory of behavioural modifiability in which both environment and heredity play a part, he contends that the child and the environment condition each other; he rejects both maturation and critical periods and in particular he objects to the use of age as a variable, preferring to see processes occurring along a time continuum in terms of a sequence of environmental experiences. There are two difficulties with this approach: competent analysis of his hypotheses would demand the use of the Markov process, a promising but as yet only partially developed model within the behavioural sciences; secondly, the multivariate nature of most developmental variables raises considerable questions about any approach based on a series of simple interactions between organism and environment.

The well known problem raised by Tanner (1962, 1963) about the need to study individual growth curves rather than attempting to analyse grouped longitudinal data in terms of chronological age is an important issue, especially as it may be argued that young children who are starting school are in some respects at the point of a 'critical cognitive barrier' such as those experienced in the physical growth of older children. However, as with the problems raised by Gewirtz, Tanner's approach does not fit easily into multivariate situations; statistical models have yet to be developed to incorporate unique phases of rapid development and stationary development within a larger pattern of multiple relationships.

A related difficulty in interpreting development along an age continuum is shown up in a series of studies by Hindley and Moore (inter alia Moore, Hindley and Falkner, 1954, Hindley 1965 and Hindley and Owen, 1978) in which a considerable number of psychological and physiological variables are measured on a population of 200 children from birth to maturity. Among the findings from these studies is the evidence of major fluctuations in individual patterns of development over time. Such differences clearly limit the size of relationships that might be discovered across the sample as a whole.

Nevertheless, it may be argued that despite its conceptual limitations, age still serves as a crude and usable approximation of developmental status in the child, reflecting at least the time span during which the child's cognitive functions have been subjected to environmental influences and to the associated physiological maturation processes (the latter affecting the degree to which the environment can alter the functions themselves).

The time variable, although manifesting itself more clearly than does age, is an even more difficult concept to build into longitudinal analysis. Campbell and Stanley (1963) point out that "experiments we do today will need replication and cross-validation at other times and under other conditions before they can be theoretically interpreted with confidence".

This author has elsewhere pointed to the range of interpretation possible for time: time measured as a strictly interval scale; seen as history or as elapsed time; assessed as personal time in the form of rhythms and cycles; time as perceived by the observer; time viewed simply as temporal sequence; time as subsumed within the experienced environment and within the concept of age, where it is of particular relevance to longitudinal analysis; and time as it is interwoven into the study of child development and the complex of interacting maturational and environmental variables. It is the second last form that is of particular interest in this research.

Time is of course closely bound up with age. Schaie (1970, 1972) has written extensively on the relationship between time, age and cohort. He has developed an hypothesis in which he claims that a three-factor model (age, cohort and time of measurement) is needed to make parameter estimates for the population of interest. He recognises that the age of the sample can always be inferred from cohort and time data, but considers that all three variables need to be identified separately for a fuller understanding of development. Although this view has been challenged (Baltes, 1968, for example), the complexity of examining development over both age and time is evidenced by the debate on the issue.

The use of age and time in the present research needs brief specification. As Simon (1971) points out, chronological age is to some extent a surrogate for

physiological and sociological age. In this research it may be seen as a surrogate for physiological and cognitive age, alongside the more sensitive measures of parental environment and child ability. In the same way time may be seen, within the context of the child's experience of school environments during the research period, as a measure of the length of such experiences, measured separately for the nursery and reception classes.

Given the decision to take account of these variables, the question arises of how they are to be incorporated in the analyses.

#### 5.171 Use of age variables

There are various ways in which age could be employed. If it were to be used as a covariate, either for the criterion variables (post-test attainment) or as a variable for which the dependent variable would need to be adjusted, it would mean that the size of the 'age effect' could no longer be assessed alongside the effects of other variables in the model; it would also mean that a fairly sizable amount of shared variance - criterion variance in common between age and a number of other variables - would be prematurely removed from the model. The fact that there are various age variables is an additional reason for not attempting any such artificial correction of the data.

Accepting the need to treat the age variables as independent variables alongside the other predictor measures, the question then arises as to whether age should be entered into the path analyses as a causally prior variable (within the restricted context in which the term causal is used throughout this research). In one sense it can be seen as the earliest of all the indicators; on the other hand its current effects are coterminous with those of many other variables. A child's physiological or cognitive 'age' expresses the degree to which the average sample child of that age would be able to cope with a particular task, alongside other determinants of performance. The fact that separate ages are taken into account for each of the three main sets of child measures is a recognition of the fact that within the research context each child has its own set of test ages, not necessarily the same as those of other children. It is at all times a competing rather than a prior variable.

The decision to take account of each testing age as a separate variable added considerably to the flexibility with which the relatively large research project could be handled by the single E. As an additional refinement each age was measured to the nearest half month to add to its potential accuracy as a predictor.

5.172 Use of time variables

The use of time was more restricted in the present research. As previously explained, the length of programme time for different parents was measured in a more sensitive way than simply by assessing the number of weeks spent on the programmes. (For example, partial or late attendances were credited with less points than were full attendances.) The main value of time was seen in its potential for assessing the relative effects of the length of nursery and reception class experience on the child's final levels of reading and mathematics.

This aspect of the research has to be qualified strongly by the caveat that a great many of the children were still at the age where the schools were only starting to introduce them to the concepts of reading and mathematics. Thus it was not expected that the size of these time effects would be large or even statistically significant. It was hoped however that what effects might be found could point to differences between the nursery and reception class experiences, in relation to ultimate academic attainment in the two basic skills.

It was decided to create two time variables: time in nursery, with the raw variable measured from the date of the nursery attainment test to the end of the child's stay in the nursery class; and time in reception, with the raw variable here measuring the period from the entry into reception class up to the date of the post-tests. There would clearly be an inverse relationship between these two variables, although for reasons of the variation in test schedules the combination of raw time in nursery plus raw time in reception would not be identical for each child.

The difficulty with these variables was that both of them contained a large component of cohort age. In other words, a child who started nursery class earlier than most of the sample children, for example, would have a higher cohort age and thus spend relatively less time in the nursery class and relatively more time in the reception class. In such a case it was essential to take account of cohort age if the variables' value as measures of school experience were to be assessed as contributing to performance alongside or in addition to the contribution of cohort age.

The means of taking account of cohort age in this situation caused considerable problems and a variety of methods were examined. The simplest approach was that of subtracting cohort age from time in nursery and likewise removing cohort age from time in reception. Such an approach suffers from all the difficulties of change scores, including problems of differing variances. Other approaches were considered, including that of standardisation of the variables, but then the problem arose of the interpretation of the final products.



It was decided to attempt to regress post-test attainment on both time in nursery and cohort age as separate predictors, and likewise to regress attainment on time in reception and cohort age as predictors. It was hoped that separate latent variables could then be constructed (for nursery and reception), representing a composite of both time and cohort age in each case. This procedure was attempted, including variants in which the values of time and age for each cohort were 'school-centred', in order to take account of the slight differences in school practice over the age of transfer from nursery to reception.

Use of the two latent variables in the preliminary path models showed, however, that both were so heavily weighted by cohort age that they lost their value as variables identifiably distinct from test age.

The final resolution of this problem, involving the school-centred standardisation of the nursery and reception time periods and of the cohort age, and the subtraction of standardised cohort age from each of the standardised time variables, is integral to the results described in the analysis in chapter 6.00. Further discussion of the issue is held over until then.

#### 5.173 Variables available for analysis

To summarise, the following age and time-related variables were obtained or derived for the analysis:

Age at nursery testing (all age and time variables measured to the nearest half month)

Age at reception class testing

Age at post-testing

Cohort age = (child's birth date - 1.1.1972)

Time in nursery ( = date of completing nursery period - date of nursery tests)

Time in reception ( = date of post-tests - date of commencing reception class)

Reading meetings attended (totals scored to nearest quarter of a meeting attended\*, multiplied by 4)

Reading meetings weighed<sup>t</sup><sub>h</sub> (weighting in terms of outside distractions at venue, with 'normal' weighting being 4)

Mathematics meetings attended (as above)

Mathematics meetings weighted (as above)

\* This scoring takes account of partial attendances, when parents missed a part of the meeting by arriving late or leaving early.

## 5.20 Multiple regression

Multiple regression is a direct, powerful and flexible method for analysing data sets, in particular for examining the relationships between a dependent variable and a set of predictor variables. For reasons set out in preceding subsections this is the preferred method for analysis of the fairly large set of variables assembled in the present study.

Within the range of techniques covered by the multiple regression procedure there are a number of interesting variants. Two of these involve the extraction of principal components (Theil, 1971, for example) or factor scores (Lawley and Maxwell, 1973) from a set of predictor variables, and using the derived variables as the predictors. The advantages of orthogonality and the other simplifications resulting from the manipulation of the original predictor variables could not, however, outweigh the loss of variable identity in the present study.

The theory and potential of multiple regression (MRG) are described in a great many text books. Among the best standard works on this technique are Kerlinger and Pedhazur (1973), Draper and Smith (1966), Cohen and Cohen (1975), Namboordiri et al (1975) and Theil (1971).

Kerlinger and Pedhazur in particular show the very considerable scope of MRG in handling a wide variety of statistical analyses in education, psychology, sociology and other behavioural disciplines. The technique is sufficiently robust to cope with non-normality, non-equality of intervals and even categorical differences within predictor variables. The authors also show how the set of residuals from a prediction equation can serve as an explanatory variable in its own right. Draper and Smith (ibid) present a theoretical review of regression, its strengths and weaknesses, and focus in particular on a close graphical examination of residuals to determine the effectiveness of the original regression model; the difficulty with this bivariate (or trivariate) approach is that it is of limited use in most multivariate situations.

A highly comprehensive theoretical treatment of MRG is offered by Cohen and Cohen (ibid). Although shortcomings in their review have been strongly criticised by Goldstein (1976b), they succeed in showing the technique's value in assessing unique variance contributions and in defining and testing alternative causal systems or theories. The authors emphasise the need to disattenuate the correlations in a matrix to take account of the varying levels of unreliability in the equation's variables. It is noteworthy that they examine also the power of the F-tests used to assess alternative equations. Of the other studies, that of Namboordiri et al (ibid) describes MRG as "the most versatile and powerful of the

inferential frameworks". They show its application to a wide variety of linear models, including complex analysis of variance.

These and other broad reviews of the considerable scope of MRG also point to its inevitable limitations, the most serious being the instability of its regression coefficients, its capitalisation on error in the overall prediction of outcome variance, and the difficulty of establishing the best model in a highly multivariate situation. While Kendall (1975) describes some of these limitations it is Gordon (1968) whose precise and thoughtful study of the behaviour of the inverted matrix of correlations (between the predictor variables) shows why minor variations in relationships can have an exaggerated influence on the resulting regression coefficients. Another critique is offered by Bibby (1977) who summarises the considerations which may invalidate the assumptions required for the application of least squares in a regression situation; he considers that these invalidating factors can be found in most regression analyses.

#### 5.201 Practical applications

The practical applications of MRG analysis have been widespread, though the models set up often tend to oversimplify the situation being examined, with only a few of the major predictive variables being included in the research design.

Wold (1961) presented to the Fourth Berkeley Symposium - a major milestone in the development of mathematical statistics - the theoretical rationale for applying MRG to both recursive and interdependent models of relationships, and discussed the problems and prospects of generalising to causal conclusions.

There have been a number of major and numerous minor applications of MRG to causal models in recent decades. Peaker (1967b), for example, uses stepwise regression on a large body of variables from the Plowden survey data and concludes that parental attitudes count for more of the variation in children's school achievement than do either the variation in home circumstances or the variation in schools. Goldstein (1977) considers however that the use of an extremely large number of variables in the initial Plowden regressions and the averaging of scores within schools, among other techniques whose shortcomings were not acknowledged, led to the above questionable conclusions.

The debate over the acceptability of regression-based survey analysis has been carried on far more intensively in the United States. Coleman (1966) undertook a massive study of 600,000 American school children and used regression analysis to point to the key school variables which influenced performance and contributed to inequality of educational opportunity and performance. Among the large number of critics and commentators on the findings of the Coleman Commission, Cain and Watts (1970) offer substantial criticisms, claiming that the theoretical

rationale for the selection of explanatory variables and for their inclusion or exclusion from particular models was not presented by Coleman; furthermore Coleman presents the contribution of individual variables to  $R^2$  but does not attempt to translate the findings into policy options. In reply Coleman (1970) points out that the Commission used regressions on subsets of conceptually similar variables to derive linear combinations for use in the final equations. Jencks (1972), using different statistical techniques and predicting to lifetime outcomes such as jobs, rejected Coleman's conclusions about the relatively greater importance of school variables compared to home variables.

Arguments such as these do not merit a fuller discussion here. But they indicate the flavour of the debates which revolve around the interpretation of regression models. Coleman himself (1975) re-analyses the International Education Association studies which compare the strength of school and home variables and challenges the IEA conclusions - which are based on cross-national surveys of school attainment. A paper of major importance, by Campbell and Erlebacher (1970), points out that the misuse of covariance in the regression analyses of Headstart data could even make compensatory education look harmful. A variety of other papers referred to elsewhere in the present study also base their arguments on differing interpretations of MRG analyses.

The conclusions to be drawn from such debates is not that the analyst should for safety turn to the ANOVA criteria of the significance of differences between groups but rather that the considerable potential and flexibility of MRG also implies a greater openness to misinterpretation and uninformed misuse.

#### 5.202 Particular problems

Rather than attempt a review of the many conceptual and statistical problems which can arise in the use of MRG, attention will be given here to the most troublesome of these problems as they might affect the analyses planned for this study.

The major statistical problem of coefficient instability, deriving from multicollinearity in the matrix of predictors, is dealt with at length in sections 5.21 and 5.22. The problem of establishing the best model in a multivariate situation is also referred to in those sections and again in section 5.30, on path analysis.

The further problem of the least squares capitalisation on error (when amassing predicted variance in a regression equation) is examined in detail in a previous study (Barker, 1976). In this study the McNemar (1969) shrinkage formula for correcting  $R^2$  is compared with a more conservative formula suggested by Kerlinger and Pedhazur (1973); the legitimate theoretical justification for the McNemar formula as compared with the ad hoc Kerlinger and Pedhazur solution, as well as a comparison of the results of applying each formula, are presented by

Barker; the comparison suggested that the McNemar formula could be used safely. This formula is used throughout the present study to correct all derived  $R^2$  parameters; it is given by

$$\bar{R}^2 = 1 - (1 - R^2)((N - 1)/(N - k))$$

where  $R^2$  is the multiple prediction or variance accounted for  
 $\bar{R}^2$  is the corrected or shrunken  $R^2$   
 $N$  is the number of cases, and  
 $k$  is the number of predictors.

There are there are three further important conceptual problems in MRG which need to be considered.

The stepwise procedure is one of these. Despite its widespread usage in many large scale studies there are several fundamental objections to it, on both conceptual and statistical grounds.

Firstly, a detailed examination of the error structure and the sensitivity of the derived coefficients in the face of any collinearity, serious or otherwise, in the predictor matrix of a multiple regression equation

$$B = [X'X]^{-1}X'y$$

where  $B$  is the vector of regression coefficients  
 $y$  is the predicted vector, and  
 $X$  is the set of independent variable vectors,

means that any stepwise inclusion or exclusion is absolutely dependent on the error structure in the matrix at that particular stage. Without a detailed examination of each matrix as it is formed and inverted, it is impossible to be sure whether a stepwise 'decision' by the computer is legitimate or fortuitous.

A further statistical consideration is that forward and backward stepwise procedures can yield differing results regarding inclusion and exclusion; thus the principle of consistency cannot be guaranteed in a situation where there are no statistical grounds for opting for either approach.

The most fundamental argument against the use of stepwise is a conceptual one. Despite the claims that a particular variable in a multiple regression situation (as opposed to a path analysis situation) has some conceptual priority, it is invariably possible to show either that there is no legitimate order of priority in a set of competing predictors, or alternatively where there is, that the model specification should be that of path analysis rather than of a single MRG equation.

Many analyses have been based on a regression model in which, say, 'intelligence', together with a motivational variable, social class and some school variable are used to predict jointly to an educational outcome. When these variables

are entered together into a stepwise programme the 'intelligence' variable is nearly always the largest contributor and the first to be taken up in the equation; social class may well follow; motivation might just be accepted into the equation according to the F-test criterion for the difference between the two-variable and three-variable predictions, but the school variable could well be left out. A minor change in the error structure could easily reverse this 'decision' on entry.

When all variables are entered simultaneously the criterion for acceptability or otherwise is the probability of the derived coefficient for a particular variable, with unique variance as a further though more fallible criterion. These in combination are intuitively more reasonable criteria and leave the decision to the analyst as to inclusion or exclusion, rather than to the particular stepwise technique programmed into the computer.

The conceptual question of variable priority is always a most difficult one to resolve. While it may be claimed that a child's cognitive levels inhere mainly in the child and thus have a kind of natural priority of predictive effect, in fact the child's current cognitive levels may be interacting with his or her motivational level and with certain school variables, to produce a particular level of academic attainment. On the other hand social class may well be an antecedent variable predicting to all three of the other named predictors and having no direct effect on attainment, justifying perhaps a recursive path model. Stepwise methods ignore the subtlety of relationships and deprive the analyst of the basic responsibility for deciding on the nature of the hypothesised model.

Kerlinger and Pedhazur (ibid) and many other textbooks rely heavily on the stepwise approach. Yet as Kendall (1975) shows, not only may forward and backward methods yield different answers, but even if they do that answer may not be optimal.

A second conceptual problem in MRG is that of outliers. Section 5.103 has already dealt with this issue. Its seriousness for multiple regression equations lies in the fact that a distant outlier may not only affect the variance but may have a highly magnified effect on the inverted predictor matrix and the resulting regression coefficients, as Gordon (1968) has shown. This problem has no resolution within the present study (and fortunately did not appear to arise to any important degree).

The third conceptual problem is yet more intractable. A body of theory has developed around the concept of 'suppressor variables', predictors whose regression relationship with the dependent variable (D.V.) is in the opposite direction to its simple correlation with that variable. It has been argued forcefully by Cohen and Cohen (1975), Darlington (1968), Van de Geer (1971) and others that such variables, when they are noted, should not be eliminated or explained away as a perversity. They contend that a suppressor variable, by virtue of its

particular relationships with the D.V. and with other predictors may well be 'suppressing' that part of the variance of another predictor which is irrelevant to the D.V., despite the apparent direct contribution of that predictor. Cohen and Cohen warn that it is possible to misinterpret this effect; it may be due to any one of three differing types of relationships - none of which occur frequently.

The difficulty with this interpretation is that the major review carried out by the present author into the shortcomings of the least squares algorithm and its derived regression coefficients shows that when ridge regression techniques are applied, some of the more fallible coefficients can easily reverse polarity as 'ridge k' is increased in magnitude (see section 5.22), suggesting that the original polarity may have been heavily influenced by the error structure within the matrix rather than by any abstruse 'suppression' function.

During the course of the analysis it became necessary to decide how to handle a few situations where a coefficient proved to be opposite in sign to that of the relevant correlation coefficient. In view of Cohen and Cohen's view that suppression effects did not occur frequently, coupled with doubts as to the credibility of the derived coefficients (i.e. the possibility that a sufficiently high ridge k might well reverse the polarity of an apparently anomalous coefficient), it was decided to ignore any variable within a particular equation which made only a minor regression contribution of opposite polarity to its correlation with the D.V., but to consider carefully the situation of a variable which showed its strength and reverse polarity in the V-ridge model developed in this study as an alternative to least squares regression.

### 5.203 Cross-validation

Given the considerable evidence about the fallibility of the statistics derived from regression or any other analytical techniques, it is surprising how seldom cross-validation techniques are used to test the obtained solutions. An important study by Stone (1974) set out the rationale and theoretical justification for this procedure. But long before this Campbell and Stanley (1963) and others had stressed the need to establish the validity of any hypothesised model of relationships. Within the regression model there is clearly a case for assessing the ability of a set of coefficients to fit a sample drawn from the same population but different from that on which the original set of coefficients was derived.

Among the relatively few studies in which cross-validation has been used are those of Novick et al (1972), who estimated and compared multiple regression solutions across a number of groups, and Golub et al (1979) who used the proced-

ure to indicate the preferable solution in a ridge regression model.

While it would not be feasible to carry out cross-validation runs for each of the numerous regressions to be undertaken in the present study, the establishment of a routine cross-validation procedure (described in detail in section 5.23) within the regression programme written for the study, will enable frequent checks on the validity of the derived parameters, particularly in the final equation derived for each path model.



### 5.21 The limitations of Ordinary Least Squares

Interpretations of data depend largely on the analyses which can be performed on them. There is much evidence that in multiple regression based on ordinary least squares the coefficients are unreliable for moderate or small data sets with correlated regressor variables. The effect of this limitation on what should be a powerful method for interpreting multivariate models is that such models have to be pared down not merely to essentials but to a simplistic level which does injustice to the theoretical concepts underlying the research.

There are differing approaches to the question of multicollinearity. Mosteller and Tukey (1977) see correlated regressor variables as often "measuring essentially the same quantity under different names, .... try(ing) to treat one piece of information as though it were several pieces"; this leads to arbitrariness about the allocation of the weights to be given to the several pieces. The authors consider that this situation arises very readily in social science and economics, "when we have many variables that may enter a regression relation and clumps of these variables measure much the same thing".

An alternative view which might be offered here is that while the correlations between variables suggest that many life processes are related - motivation, cognitive functioning and the quality of educational input, for example - there are also important differences which need separate identification and which make separate though overlapping contributions to some outcome variable such as performance.

Thus in psychology, sociology and econometrics almost any individual or group performance is the product of a multiplicity of variables, many of which are closely related and yet different. For example, the bright child is more likely to enjoy good health, to be better motivated, to come from an experientially richer home environment, to have benefited from better quality education, and the like. Research on a particular hypothesis or on a general model usually requires the simultaneous consideration of a reasonable number of interrelated variables, but this is difficult when using ordinary least squares (OLS) regression, and it is here that the apparent arbitrariness arises to which Mosteller and Tukey (ibid) refer.

Sample size defines to some extent the statistical limitations of OLS regression. With a sample of several thousand the standard errors of the coefficients may be small enough to offer some reasonable confidence in the

results, although even large samples have problems of over-capitalisation on certain features in the regressor matrix. For smaller samples the multiplicity of causation is recognised but statistical limitations prevent adequate assessment of this multiplicity. In models based on path analysis, for example, regression coefficients play a major part; yet the conclusions are seldom confirmed in replication studies. Other small sample models make use of time series analysis; this again is rooted in the OLS algorithm and is thus equally subject to the shortcomings described above.

#### 5.211 Statistical limitations of the regression method

The inadequacy of OLS regression coefficients is shown in several studies reviewed by Beaton et al (1976). The unreliability of the coefficients - differences not only in the first significant digit but even in the sign of the coefficients when various kinds of regression programme are applied to the same package of data - are due not only to shortcomings of the programmes but to the OLS method itself. The authors find that programmes using 40 decimal digits of accuracy bring virtually no improvement in regression results; coefficients fluctuate wildly as a result of seemingly minor errors in the data as well as the choice of algorithm, package programme, etc. When the analysis with three variables is extended to a 6-variable solution the coefficients differ in some cases by orders of magnitude. Deegan (1976) finds that multicollinearity is the most important methodological problem leading to model misspecification in regression analysis. As Cooley and Lohnes (1971) point out, the presence of sample variation means that very little reliance can be placed on individual coefficients when there is substantial collinearity in the regressor variables; one study they cite showed that for samples of less than 200, simple correlations were better predictors than regression weights in the majority of cases examined.

The reasons for the unreliability of the OLS coefficients are to be found in a combination of the multicollinearity between the regressor variables and the random differences between similar samples.

Gordon (1968) presents an unusual analysis of the aberrant behaviour of the coefficients. He shows how minor changes in relationships between correlated regressors can lead to major changes in the values and even in the signs of the regression coefficients. He analyses features of the inverse regressor matrix which contribute to this phenomenon. While some of the problems cited by Gordon such as the different repetitiveness in differing sets of variables are clearly conceptual problems to be dealt with by the researchers, his analysis highlights the fact that random errors which result in modest differences in

correlations between regressor variables can lead to a major 'tipping effect' on the model, changing the regression coefficients radically for reasons unrelated to any conceptual hypotheses.

More general shortcomings in the application of the Gauss-Markov theorem are reviewed by Bibby (1977). He cites nine assumptions which need to be met before acceptance of the ordinary least squares estimator in a linear model. Many of these assumptions are not tenable in practice and this fact may contribute to the grave limitations of the OLS estimator. Bibby contends that if it is not necessary for the estimator to be unbiased, then even under normality these estimators can be improved using 'Stein' procedures and the ridge estimators of Hoerl and Kennard (1970a, 1970b).

#### 5.212 Attempts to overcome the problem

There have been many different attempts to overcome the problem of the unreliability of OLS estimators. One major approach has been that of stepwise or hierarchical regression. In the former, successive regression equations determine the sequential entry (or exit) of the most (or least) 'important' predictor within each successive set of possible (or already included) variables. In hierarchical regression the researcher decides on the order of entry of variables or sets of variables into the equation and abides by the coefficients as yielded by this method. Cohen and Cohen (1975) and Kerlinger and Pedhazur (1973) give examples of these and other related methods.

Such methods involve major conceptual assumptions. In stepwise regression the assumption often made is that the regression equation will itself yield the evidence on the order of importance of the regressor variables. Yet all the evidence on coefficient instability indicates that this is a chance process, with a particular combination of sample idiosyncracies leading to one or other coefficient being thrown up as the most important, often in such an exaggerated form as to outweigh the cautions of the given standard error. Even with very large samples there are major questions about the wisdom of leaving a statistical technique to determine the relative importance of predictors in terms of stepwise procedures. The naive statement, often presented in research reports, that variable B contributes 40 per cent of the criterion variance, and the remaining four variables a further 20 per cent of the variance, as 'shown' by the stepwise procedure, ignores the fact that in reality the first variable usually shares a great deal of the variance with the other predictors; moreover, if that ostensibly major variable were removed from the model the remaining variables would probably account for most of the previously explained variance in many situations of multicollinearity.

The hierarchical procedure does restore the power of decision to the researcher on the basis of a conceptual stance and the statistical evidence, within the limitations of coefficient instability. But even if there were clear conceptual grounds for the prior entry of a certain variable or subset into the regression, there is no reason to suppose that the lion's share of the variance which this initial entry then takes is a justifiable share. The problem still remains that much of that variance is in reality shared with variables or subsets yet to be entered.

The wide scope afforded by employing regression procedures in line with ones prior conceptual assumptions was shown by two recent reports. Broman et al (1975) examined a large number of variables derived from 12,000 White and 14,000 Black American children; the authors' prediction of IQ at 4 is based on stepwise regression procedures which suggest that for each ethnic group the prenatal variables account for most of the variance and the neonatal and infancy-childhood variables for very little. As Tizard and Plewis (1977) point out, the statistical limitations of this study encourage the dubious inference that genetic rather than environmental factors are of overwhelming importance in determining early I.Q.

In contrast, Smilansky et al (1976) have used a combination of hierarchical and stepwise regression methods in a longitudinal study of over 300 Israeli infants of Moroccan and European origin. In the prediction of IQ at 4 the environmental variables are entered in a stepwise regression, but parental education, sex and ethnicity are hierarchically forced in last. The result is that parental education adds only 3.6 per cent to the 34.8 per cent variance already explained, achieving a highly significant but negative regression weight, sex is totally removed from the regression and ethnicity yields a non-significant weight. What is in other ways a most useful study is damaged by the inadequacy of the final analyses.

The second major alternative to reliance on ordinary multiple regression coefficients is the orthogonalisation of the predictor variables, either in the form of principal components or by the use of one or other kind of factor analysis to derive a set of factor scores which are then interpreted for their conceptual meaning and used to derive regression coefficients. It is clear that this method eliminates the uniqueness of each variable, so that one can only judge sets of variables as a whole in their ability to offer a combined prediction. Lawley and Maxwell (1975) present the case for the use of factor analysis in a maximum likelihood solution of the regression situation. They offer a chi-squared criterion for limiting the number of factors and show how it is possible to obtain unbiased estimates of the regression weights and multiple

correlation by taking account of error in the measurement of the independent variables.

Apart from the loss of the identity of the separate variables in the orthogonal component solution there is also the fundamental disadvantage, pointed out by Van de Geer (1971), that such solutions are highly dependent on the specific variables chosen in any one study. A particular solution can be completely altered by the addition or removal of a single variable.

In addition to these two major alternatives to the ordinary process of multiple regression there are at least four other alternatives to be found in the literature. The simplest of these is the unitisation of regression weights. Both Schmidt (1971,1972) and Wainer (1976) present what may be termed the despairing solution that is forced on psychometricians by the instability of the regression coefficient, namely to treat all the variables in the model as of equal importance, with weights of unity.

A more serious approach is offered by Gorsuch (1973) who suggests some interesting methods for interpreting data in terms of the division of variance. One of his methods attributes variance overlap to specified variables in terms of a sequence of causal hypotheses, relying on triangular decomposition to yield the remaining matrix after the removal of each predictor. Another method is to separate the variance into overlapping and non-overlapping components, creating new variables to represent overlapping components in different sectors of the overlap.

Studies making a sophisticated use of basic indices have been offered by Lunzer et al (1976) and Stevenson et al (1976). Both studies rely to a considerable extent on ingenious development of correlational evidence which, whatever its confirmatory virtues, is at best a kind of repetitive single variable explanation for the prediction, ignoring as it must the contribution of similar variables to the particular prediction.

One further alternative is the jack-knifing procedure, using various statistical criteria for the removal of outliers, thus reducing this form of error within a particular data set; or alternatively testing the equation by taking out each case in turn. There are difficult statistical and conceptual issues raised by the criteria applied within any one form of jack-knifing, in particular those methods requiring the removal of outliers from what is assumed to be an ordinary random sample.

5.22      V-ridge regression: theory and evidence

The previous two sections have dealt with the main problems encountered when using multiple regression; these problems are associated with the use of the ordinary least squares algorithm - based on the well known Gauss-Markov theorem - for solving regression equations.

It is possible to argue that even somewhat problematical procedures such as the automatic stepwise regression programmes provided by various computer packages may be a response to the unreliability of the least squares coefficients rather than any deliberate decision to place responsibility for coefficient selection in the hands of the computer instead of the analyst.

Although multiple regression has been used a fair amount in a variety of situations it is not in general a popular first choice of analysts. Yet, as already pointed out, it has much to offer in the very direct insights it gives into predictive models in which a number of independent variables are in competition.

Preliminary trials with the pre-test and mid-test data in the present study as well as earlier multiple regression work in a previous study (Barker, 1976) showed that the limitations in the reliability of the coefficients were such that a satisfactory alternative would have to be found or developed if the results of the intervention programme were to be fully credible. Although any unreliability would only become apparent if or when cross-validation checks were carried out on the various estimators, the presentation of a series of sets of least squares coefficients as final evidence on the different models was not considered satisfactory.

An examination of the many alternatives described in the statistical literature provided a number of references to a relatively new but little known procedure called ridge regression. It appeared that this particular estimator has an unusual potential for overcoming the main limitations of ordinary least squares. However even within a fairly large selection of contributions on a variety of ridge methods and their application, it was apparent that most of them are highly subjective, with only a few offering any kind of objective or non-stochastic ridge solutions.

The following sub-sections describe the theoretical underpinning for the method which was eventually developed for use in the study.

## 5.221 Ridge regression: its origins

Examining first the justification for a completely different form of regression, known as ridge, this is perhaps the point at which to describe its features and brief history. Hoerl (1962) first offered the concept of ridge analysis in a little noticed contribution. Eight years later Hoerl and Kennard (1970a, 1970b) described a method whereby the insertion of a constant  $k$  into the diagonal of the  $X^tX$  regressor matrix has the effect of reducing the extreme behaviour of the regression coefficients based on that matrix; they saw it as a means of making the system

$$\left[ X^tX + kI \right] B^* = X^ty$$

behave more like an orthogonal system, with  $B^*$  as the ridge regression vector.

The authors discuss the properties of the ordinary best linear estimator in regression situations; the OLS method yields a squared distance between the estimated regression coefficient  $\hat{B}$  and the true coefficient  $B$  which can be defined in terms of the model's error variance  $\sigma^2$  and the trace of  $X^tX$ ,

$$\text{Var} \left[ (\hat{B} - B)(\hat{B} - B) \right] = 2\sigma^4 \text{trace} \left[ X^tX \right]^{-2},$$

when the error is normal.

This is the same as

$$2\sigma^4 \sum \left( \frac{1}{\lambda_i} \right)^2,$$

where  $\lambda_i$  are the eigenvalues of the regressor matrix,

so that as  $X^tX$  moves from a unit matrix towards an increasingly correlated ill-conditioned matrix, the estimated coefficients are clearly pushed into error at an accelerated rate.

The coefficients derived from the ridge estimates, namely  $B^*$ , are then described. For  $k$  not zero  $B^*$  is shorter than  $B$ ; moreover, as  $k$  approaches infinity, ridge  $B^*$  is slowly reducing to zero. Hoerl and Kennard go on to define the ridge trace as the graphical plot of the set of  $B^*$  for each corresponding  $k$ , and prove that a value of  $B^*$  can be found which has a minimum length subject to the condition that the residual sum of squares is itself a minimum.

For any estimator  $B$  the residual sum of squares can be written as

$$\begin{aligned} \phi &= (y - XB)^t (y - XB) \\ &= (y - X\bar{B})^t (y - X\bar{B}) + (B - \bar{B})^t X^tX (B - \bar{B}) \end{aligned}$$

$$= \hat{\beta}_{\min} + \hat{\beta}(\bar{B}),$$

where  $\bar{B}$  is the true coefficient vector.

Using  $(1/k)$  as the Lagrangian multiplier, the authors show that the residual sum of squares in terms of  $B^*$  can be written

$$\begin{aligned} \hat{\beta}^*(k) &= (y - XB^*)^t (y - XB^*) \\ &= \hat{\beta}_{\min} + k^2 B^* (X^t X)^{-1} B^*, \end{aligned}$$

$$\text{where } B^* = (X^t X + kI)^{-1} X^t y.$$

In effect, one seeks that value of  $B^*$  which minimises the function

$$F = (y - XB)^t (y - XB) + (1/k)(B^t B - R^2)$$

where  $R^2$  is the squared length of the regression vector  $B$ .

The authors' most important finding is that of the relationship between the decrease in the total variance of  $B^*$  (and thus its mean square error) and the corresponding increase in the squared bias with the increase of the ridge constant  $k$ . It is thus possible to move towards a small positive  $k$ , take a little bias but substantially reduce the variance. A plot of the movement of the mean square error (MSE) function shows this rapid fall in the variance of  $B^*$  with the initial movement of  $k$  away from zero. The squared distance between  $B^*$  and true  $\bar{B}$ , defined by

$$\sigma^2 \sum [\bar{\lambda}_i / (\lambda_i + k)^2] + k^2 \bar{B} (X^t X + kI)^{-2} \bar{B},$$

goes through a minimum, this point moving closer to  $k = 0$  as the magnitude of the squared true regression vector  $\bar{B}^t \bar{B}$  increases.

It is clear that Hoerl and Kennard's formulation requires advance knowledge of the true regression coefficient, if  $k$  is to be accurately determined. Since the true coefficient cannot be known in advance,  $k$  has to be chosen by stochastic means and it is this feature of ridge regression which may have held back many field workers who might otherwise have used ridge in preference to OLS when dealing with practical regression problems. The authors themselves offer a set of guidelines for the choice of the most suitable  $k$ . Some of these are entirely stochastic, relying on criteria such as that "the residual sum of squares will not have been inflated to an unreasonable value"; alternatively they offer a method of deriving the optimal values of  $k$  by determining initial  $k$ 's on the basis of a canonical transformation of the vectors and using the initial regression coefficients, then inserting these  $k$ 's and continuing the process until stability is achieved. In practice the authors (1970b) rely mainly on visual inspection of the graph of the ridge trace (the plot of the  $B^*$  obtained for every particular value of  $k$ ) to determine the region in which the



system of ridge coefficients stabilises. A typical pattern is shown in figure 8, illustrating the kinds of changes that occur in the values of a set of regression coefficients as  $k$  increases.

#### 5.222 Recent developments

Following on the Hoerl and Kennard articles there have been a number of developments of the basic principle of ridge regression. Four are mentioned here as indicating the range of possibilities opened up by the pioneering contribution. One approach, that of Vinod, will be dealt with separately after a brief review of the other key issues in the field of ridge methodology.

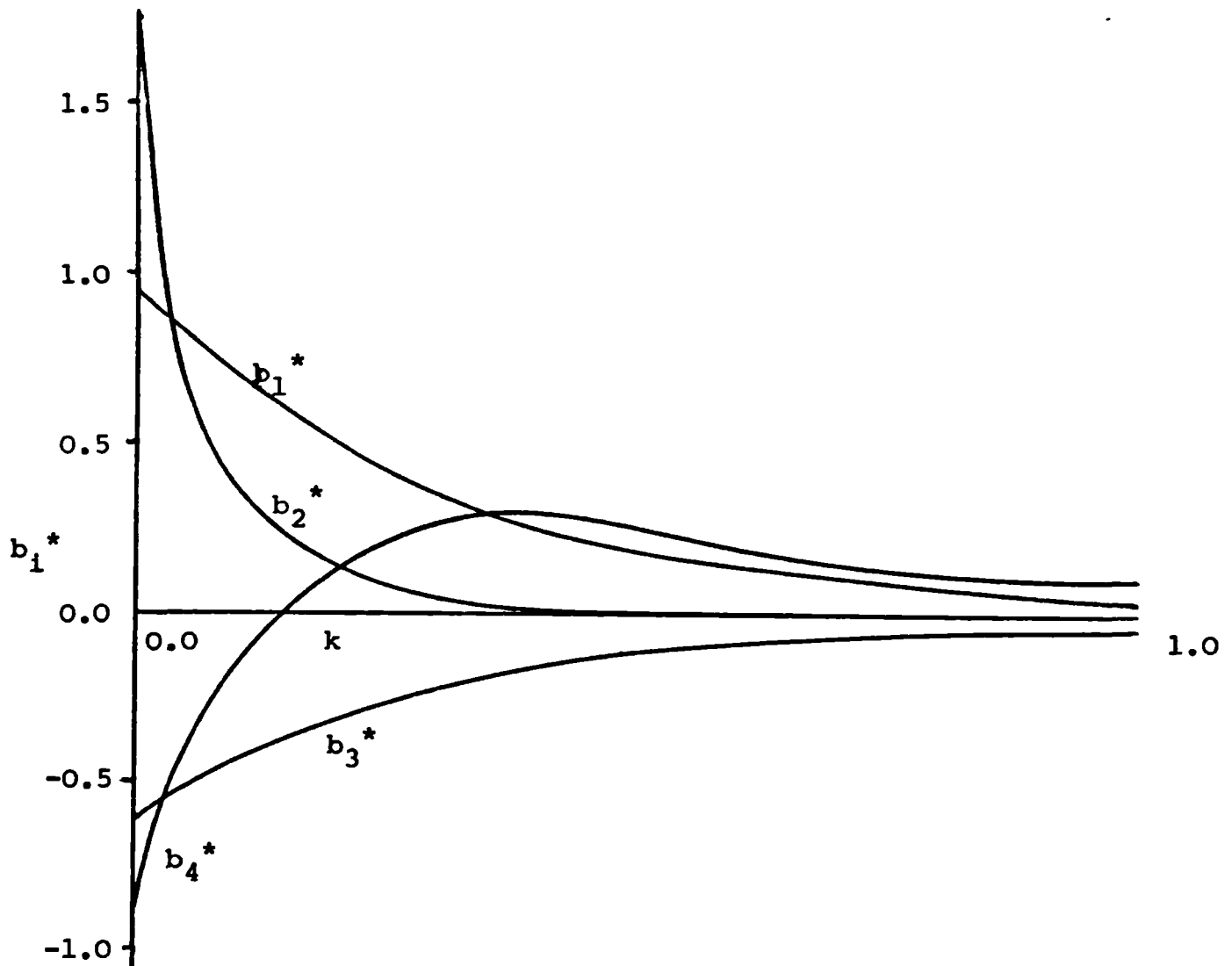
The remaining three approaches all centre on the statistical problems involved in determining a suitable  $k$ . Goldstein and Smith (1974) review the MSE properties of several different types of shrinkage estimators. The authors discuss the relative merits and disadvantages of each type of estimator and show that contrary to a suggestion by Hoerl and Kennard, a ridge coefficient based on a set of individual  $k_i$  for each element in the diagonal does offer an improved MSE performance over ridge regression based on a single  $k$ . Among other suggestions they described briefly an alternative ridge estimator (their section 4.3) which makes the analysis particularly sensitive to variation in the eigenvalue spectrum.

Obenchain (1975) looks at the ridge problem from the theoretical viewpoint of the likelihood ratio and goes on to indicate that most of the criteria which have been proposed for the ridge estimator are dependent on the criterion  $y$ . He offers a shrinkage estimator based only on the parameters of the eigenvalues and  $k_i$ , with the optimum choice of  $k_i$  depending on the sum of squares of correlations between coefficients. His choice of solutions among generalised ridge and other estimators is found in each application by monitoring a likelihood ratio statistic which Obenchain has developed.

Guilkey and Murphy (1975) present a more empirical approach. They suggest a directed ridge regression in which only those diagonal elements corresponding to relatively small eigenvalues are altered. There will be less bias in this model than in the usual ridge situation. Another empirical model is developed by Wichern and Churchill (1978) who offer a set of mechanical rules and a graphical procedure for determining the best  $k$  value.

A number of workers have looked closely at the performance of ridge estimators and other statistical techniques aimed at reducing the main problem faced in multiple regression. As well as Goldstein and Smith's review there have been several other key studies in this area. Gunst and Mason (1977)

Figure 8. The ridge trace: examples of typical alterations in coefficient values.



$b_i^*$  ridge coefficient values .

$k$  ridge constant (uniform addition to diagonal of regressor matrix)

$b_1^*$  slow movement from high value as  $k$  increases

$b_2^*$  early loss of most of its value, suggesting that it is highly related to error structure in the matrix

$b_3^*$  relative stability as  $k$  increases

$b_4^*$  alteration of sign with increasing weighting of diagonal

compare the performance of OLS, principal component, ridge regression, latent root and a shrunken estimator, when deriving regression coefficients in a simulation study. They find that all the alternatives perform better than ordinary least squares; principal components and latent root perform best overall, but the ridge estimator is seen to have the potential for a smaller MSE than have any of the other solutions. Hocking et al (1976) present a unified theoretical development of a number of different estimators and offer comparisons between the results from generalised (separate  $k_i$ ) and simple (single  $k$ ) ridge procedures.

Of the more empirical studies, Dempster et al (1977) offer an impressively wide-ranging examination; they compare the OLS solution with some 56 alternative estimators, on a range of 160 artificial data sets. Among the groups of alternatives tested are a variety of stepwise and other selection methods and variations of the principal component, Stein shrinkage and ridge regression methods. They show the situations under which the different approaches yield more appropriate results, in terms of various criteria, and conclude that ridge emerges as an important tool.

Other useful empirical studies on both simulation and practical data are those of Marquardt and Snee (1975) and of Price (1977), who offers a study in a specifically psychological context. An unusually interesting application of ridge regression to a prediction situation is described by Brown and Payne (1975) who prepared a model for forecasting the results of the British General Election in 1974. The model was based on an unusually high  $k$  figure, which the authors believe reflected the large noise to signal ratio found in pre-election regression studies. In a situation involving approximately 650 seats their election night ridge-based forecasts were correct after some 30 seats had been declared.

### 5.223 Criticisms of ridge

There have understandably been some criticisms of the concept of ridge regression. Many writers who describe its advantages are also aware of its shortcomings. Only a few workers reject it altogether. Coniffe and Stone (1973) argue that the limitations of the ridge technique are very serious because it ignores the possibilities that the ill-conditioned nature of the  $X^tX$  matrix may be due either to an inadequate assemblage of data for the model under study, or to a high interdependence of some of the variables in the model. Farebrother (1975) argues that in place of the subjective method of determining ridge  $k$  there are several minimum MSE estimators which can give consistent results.

The strongest caution expressed about ridge comes from McDonald and Galarneau (1975). They show that no consistent value of  $k$  can be certain to yield

a ridge estimator which is better than least squares for all unknown coefficient vectors, and consequently they propose a series of rules to be applied in the choice of  $k$ . They offer an algorithm based on the residual sum of squares for the OLS estimate, on the eigenvalues and on the squared lengths of the OLS and ridge coefficients. This algorithm determines the circumstances in which either the ridge coefficients or the OLS alternatives should be chosen. In particular, they show that ridge performs poorly when the criterion  $y$  is parallel to the minor principal axes of the regressor data. In the Monte Carlo simulations tested by these authors they find that the ratio of the MSE for the ridge estimators to that for the OLS estimators varies from .33 to 1.34 in the cases studied.

In particular situations ridge has been found to do well. An earlier study by McDonald and Schwing (1973) compares the performance of OLS and ridge on a data bank of mortality statistics related to pollution and other socio-ecological variables. They find serious instability in the derived OLS coefficients and conclude that the ridge coefficients, with one notable exception, appear to be reasonable values on which to base the quantification of the relationship between mortality and the hypothesised causal variables. Gibbons (1978) offers an examination of ten promising ridge algorithms - Vinod's ridge is not among this group - and concludes that three algorithms perform well overall. However she emphasises that none of the estimators was able to predict the optimal  $k$ .

Kendall (1975) dismisses the concept of stochastic ridge in a footnote to his study of multivariate analysis. He recognises that an ill-conditioned matrix of covariances can be brought nearer to stability (i.e. its determinant is further from zero) by adding a positive constant to the diagonal, the choice of constant being dependent on the behaviour of the matrix under variation in this constant. It is not plain to him that this admitted distortion of the data (thereby diminishing correlations among variables) has any theoretical justification. As with other critics, his attention is confined to the subjective Bayesian approach.

A further criticism has been offered by Rozeboom (1979) who describes ridge regression as an intriguing new toy which may some day evolve into a useful if limited tool but is still too fragile to do real work. Unfortunately his case relies entirely on proofs within a single-variable regression model, whereas the whole concept of ridge relies on the weighting of the diagonal within a multivariate, multicollinear situation, and on the effects of this weighting on the overall error structure and consequent derived set of coefficients.

A more weighty criticism is offered by Draper and Van Nostrand (1979). After dealing with the limitations of the James-Stein shrinkage estimators and

the use of cross-validation to derive other shrinkage estimators the authors review nearly all the major ridge algorithms described in the literature to date. They examine the models in turn, based as they are on the "unfortunate" fact that the optimum ridge  $k$  depends on the unknown regression parameters. The subjective use of the ridge trace for estimating  $k$  is condemned; a more sympathetic view is taken of attempts to establish a mathematical derivation of this constant. Various iterative methods are reviewed, as are methods involving the use of the eigenvalues of the predictor matrix and attempts to establish criterion judgements for choosing between OLS and ridge solutions. Different simulation studies are examined; the authors consider that many studies showing favourable outcomes for ridge, as compared with OLS or other estimators, may have been biased by the nature of the simulations; moreover there has been little application of shrinkage estimators to experimental design.

Draper and Van Nostrand conclude that there are only two sets of circumstances in which ridge regression is appropriate, both of them involving some additional external information: one is a Bayesian formulation, given a posterior mean, and the other is a situation where a spherical restriction of the least squares formulation applies, namely that  $B'B = a \text{ constant}$ , where  $B$  is the vector of regression coefficients. It is important to note that the authors' review discusses an iterative fixed point estimator of Vinod (1976a) but does not refer to the alternative non-stochastic procedure described in Vinod (1976b).

The final point made by Draper and Van Nostrand (ibid) is that the extended inference that ridge is "always" better than least squares is, typically, completely unjustified. Yet there are few if any ridge authors who would dare claim such infallibility for ridge; the burden of the argument advanced nowadays is that the Gauss-Markov theorem is no longer seen as necessarily the best solution and that in fact least squares may frequently be a less appropriate solution than ridge (or one of the other shrinkage estimators). The problems are essentially those of, firstly, establishing in which range of conditions does ridge offer a solution that is generally better than least squares, and secondly, determining the ridge  $k$  according to non-stochastic and repeatable criteria rather than relying on a Bayesian model or assuming the existence of additional external information about the regression parameters.

At this point there are some tentative conclusions which might be drawn about ridge. With the exception of the Obenchain study and Vinod's own work, every model and all the examinations of ridge performance have been based on the stochastic method of observing the ridge trace (plotting the actual coefficients against  $k$ ) or using derived coefficients in some other way to decide post hoc on the best  $k$  for the final regression. Even within this stochastic framework there is no consensus on the best criteria for determining  $k$ , nor any tested method for limiting the uncertainty which is endemic in this area of subjective decision.

It is in the context of the above doubts that the reasons may be found why, after some eleven years work by statisticians on this procedure, it remains almost unknown and seldom used in the field of practical studies.

#### 5.224 The Vinod method

One ridge method does appear to offer an answer to the concern of the research worker who wishes to find an alternative to ordinary least squares that is founded on objective criteria rather than the subjective judgement of a ridge trace. Vinod (1976b) presents a ridge regression method whose elements include a direct and non-stochastic technique for deciding on ridge  $k$ . This method is based only on the eigenvalues from the regressor matrix. Vinod offers several statistical criteria for selecting the best  $k$ . The particular criterion to be described in this paper is his 'Index of Stability of Relative Magnitudes' (ISRM) of the regression coefficients, which offers a single value for the best  $k$  on a particular regressor matrix.

Vinod discusses the well recognised features of the ridge trace, as normally portrayed in the plot of the individual coefficient values vs. ridge  $k$ . As he points out, finding a  $k$  for any specific problem has remained something of an art; moreover, the basic statistical justification for arguing that the aggregated MSE of the ridge coefficients is less than the aggregate based on the OLS estimates is itself invalid if  $k$  depends on the random variable  $y$  rather than on the non-stochastic matrix of regressors.

The author shows that these problems can be avoided by choosing a new scaling on the horizontal axis instead of  $k$ . This scaling, termed the multicollinearity allowance, is defined by

$$\begin{aligned} m &= p - \sum_i \left[ \lambda_i / (\lambda_i + k_i) \right] \\ &= p - \sum_i \delta_i \end{aligned}$$

where  $p$  is the number of regressors

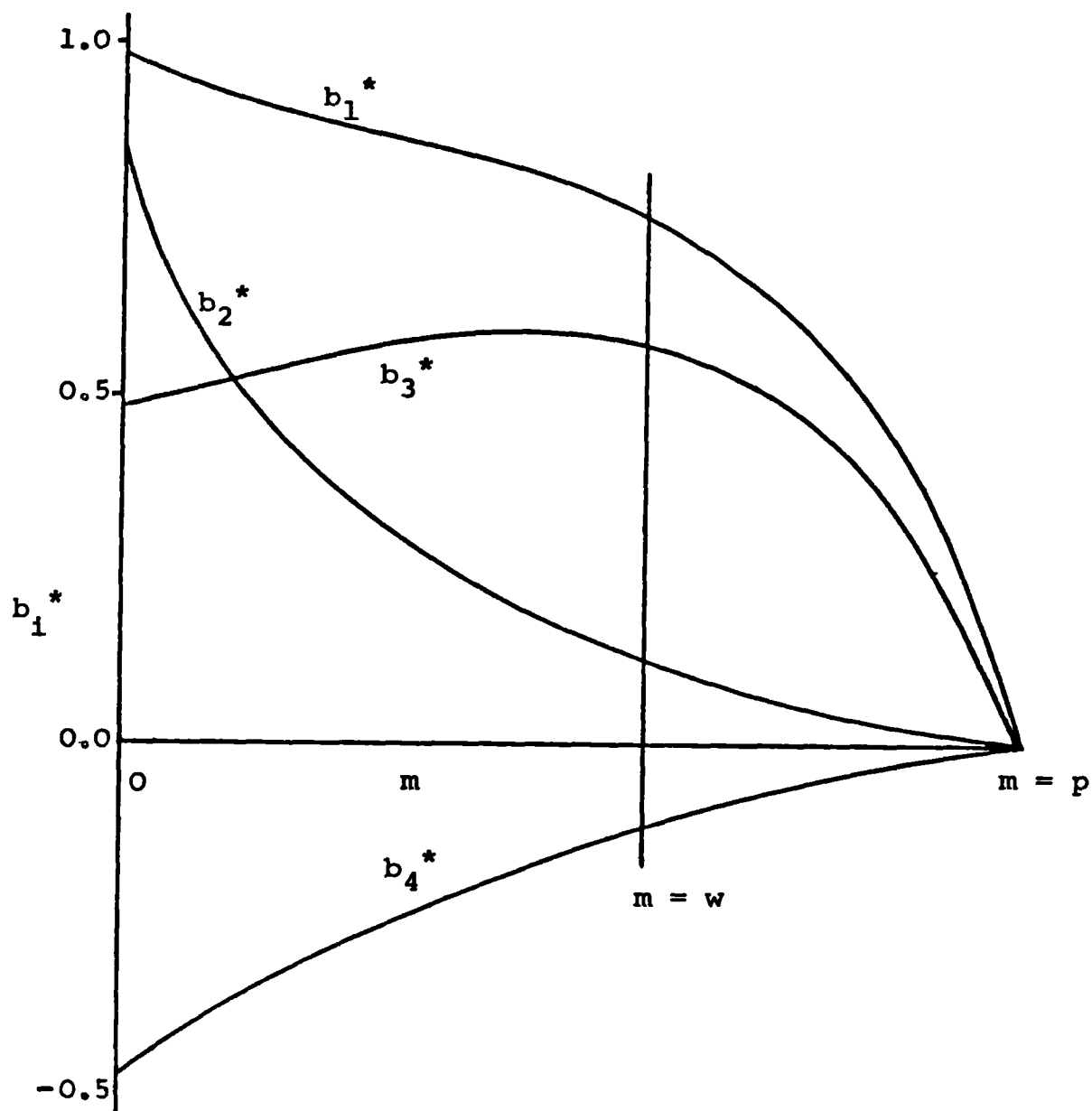
$\lambda_i$  are the eigenvalues ( $i = 1, 2, \dots, p$ )

$k_i$  are the ridge constants

In the case where a single  $k$  is used, the  $k_i$  can be replaced by  $k$  in the above expression.

A plot of each  $m$  value against the corresponding regression coefficients shows the existence of a desirable region in terms of the movement of the coefficients. Figure 9 is based on the Vinod model and illustrates the type of changes that might occur in the values of a set of regression coefficients as  $m$  increases. Whereas the graph with  $k$  on the horizontal axis (figure 8) offers no clear minima other than infinity, the graph based on  $m$  shows a limited region

Figure 9. Plot of ridge coefficients against Vinod's  $m$ .



$b_i^*$  are coefficient values

$m$  is Vinod's  $m = p - \sum_i \left[ \lambda_i / (\lambda_i + k) \right]$

where  $p$  is the number of regressors

$\lambda_i$  are the eigenvalues

$k$  is the ridge constant

$w$  is the point at which there is least movement overall in the values of coefficients  $b_i^*$

At  $m = 0$ ,  $k = 0$

of preference in terms of regression coefficients that are neither too extreme nor too reduced in size.

This leads Vinod to his Index of Stability of Relative Magnitudes (ISRM) which can be plotted for each selected value of  $m$ ; the point at which the ISRM reaches a minimum indicates the points at which the ridge system is closer to an orthogonal system than at any other point. Its chief value is that it is non-stochastic and does not depend on visual interpretation of a trace which has no clearly defined region of preference.

#### 5.225 Derivation of ISRM and its corresponding $k$

Vinod's derivation of his index reflecting the stability of the relative magnitudes of the coefficients is worth studying in some detail, as it is obviously one of the core elements if not the main key to the present author's interpretation of the Vinod method.

Vinod cites Hoerl and Kennard's evidence that the variance of the ridge estimator is

$$\text{Var} [B^*] = \sigma^2 P^t \text{Diag}(\delta_i^2 / \lambda_i) P$$

where  $\sigma^2$  is the usual error variance

$P$  is the orthogonal matrix of eigenvectors,  
such that  $X^t X = P^t \Lambda P$ ,

with  $\Lambda$  the diagonal matrix of eigenvalues of  $X^t X$

$$\delta_i = \lambda_i / (\lambda_i + k)$$

$\lambda_i$  are eigenvalues corresponding to each coefficient  
in the vector

$$\text{Since } B^* = (X^t X + kI)^{-1} X^t y, \text{ and}$$

$$\begin{aligned} \text{denoting } Z &= (X^t X + kI)^{-1} X^t X \\ &= P^t \text{Diag}(\delta_i) P, \end{aligned}$$

it is not difficult to see that  $B^* = ZB^0$ , where  $B^0$  is the unbiased OLS coefficient vector.

He goes on to show that the change in the  $B^*$  vector relative to  $k$  is

$$dB^* / dk = - P^t \text{Diag} \left[ \lambda_i / (\lambda_i + k)^2 \right] P B^0.$$

From this derivation Vinod points out that the absolute value of the change in  $B^*$  for given change in  $k$  is smaller for larger  $k$ , so that the  $k$  scale has the unfortunate property that the coefficient vector appears to become more



stable (in relation to  $k$ ) for larger values of  $k$ , even when the regressor data are completely orthogonal. In contrast, the change in  $B^*$  with  $m$  is

$$\begin{aligned} dB^* / dm &= (dB^* / dk) / (dm / dk) , \\ \text{so that } dB^* / dm &= - \left\{ P^t \text{Diag} \left[ \lambda_i / (\lambda_i + k)^2 \right] P B^0 \right\} / \bar{S} , \\ \text{where } \bar{S} &= dm / dk \\ &= \sum_i \left[ \lambda_i / (\lambda_i + k)^2 \right] . \end{aligned}$$

For orthogonal data, where  $\bar{S}$  becomes  $p / (1 + k)^2$ , it can be seen that  $dB^* / dm = - B^0 / p$ , which does not change with  $m$  nor show any apparent greater stability at larger values of  $m$ .

Having established the characteristics of the change in  $B^*$  relative to  $m$ , Vinod goes on to derive his ISRM of the  $B_i^*$ . He defines it as

$$\text{ISRM} = \sum_i \left[ (p \delta_i^2 / \bar{S} \lambda_i) - 1 \right]^2 .$$

It is possible to compute ISRM for each  $m$  figure and determine the point at which ISRM has its smallest value. This point is the one where it can be inferred that the ridge regressor system is closer to an orthogonal system than anywhere else. He sees it as a quantification of Hoerl and Kennard's concept of a stable region, with the advantage that the ISRM yields a narrow range of desirable values at which  $m$  can be chosen.

Vinod does not develop the interpretation of his ISRM statistic beyond this point, although it is evidently a statistic of unusual interest.

The ISRM successfully avoids reliance on the regression equation's error variance, so that the resulting statistic owes its existence solely to the latent roots of the regressor matrix and the particular complementary  $k$  value (or  $k$  values, in the case of  $k_i$  for generalised ridge).

Much work needs to be done to examine the mathematical behaviour of the ISRM statistic, to confirm its intuitive appeal and to delimit the situations in which it is generally applicable as well as those situations where its use is inadvisable. Vinod himself offers some preliminary cautions and algorithms designed to cope with the kind of problem cited by McDonald and Galarneau (on the poor performance of ridge when the criterion  $y$  is parallel to the minor principal axes of the regressor matrix), and to deal with the possibility that the effect of the multicollinearity in the regressor matrix may be such as to distort what Vinod calls the numerical largeness of the more significant regression coefficients, so that the less significant coefficients have larger

values than the more significant ones.

The Vinod approach could be of major importance for analysts looking for a way of minimising the serious problems inherent in ordinary least squares regression on medium to small data sets in which moderate to high error structures exist. The concepts underlying this approach are relatively straightforward and well defined; the cautions Vinod describes are important although the conditions under which they would be operative are by no means general. A great many research models are designed to focus on what are thought to be the main predictors of the criterion variable and it is less often that the problem described by McDonald and Galarneau is likely to arise.

In view of the sharp distinction that exists between the great majority of stochastic ridge methods based on visual inspection of the ridge trace and the non-stochastic determination of the Vinod ridge  $k$ , the latter method is here termed V-ridge to emphasise this fundamental difference of approach. The name is also appropriate in view of the characteristic V curve obtained from the plot of log ISRM against log  $k$  in the computer programme described in the following sections.

## 5.23 Empirical assessment...../

### 5.23 Empirical assessment of the V-ridge model

The previous pages have outlined the theoretical basis of a non-stochastic model of ridge regression derived by Vinod (1976b). The promising theoretical basis of V-ridge and in particular its potential use as an objective rather than subjective regression method led to the development by the present author of extensive computer programmes incorporating the Vinod model and offering a variety of research criteria for assessing the performance of OLS against V-ridge. The most important of these measures rely on cross-validation of coefficients derived from sub-samples of the main data set.

There are ten criteria by which the relative performance of OLS and V-ridge have been assessed in this study:

1. Proportion of total dependent variance accounted for in the main data set.
2. Residual mean square (RMS) when predicting individual scores of the dependent variable.
3. Inaccuracy in reproducing known coefficients in simulation studies.
4. t-statistics for OLS and V-ridge coefficients in the main data set.
5. Cross-validation: mean predictive efficiency over both sub-samples.
6. Cross-validation: mean increase or decrease in the RMS of V-ridge over both sub-samples, when compared to the RMS of the OLS predictions.
7. Cross-validation: coefficient stability levels across sub-samples, separately determined for OLS and V-ridge.
8. Cross-validation: maximum and minimum deviations from the mean coefficients across sub-samples, separately determined for OLS and V-ridge.
9. Cross-validation: overall stability level of V-ridge compared to OLS, based on the combined total of differences of coefficients across sub-samples.
10. Cross-validation: a combined Index of Ridge Effectiveness, based on the product of the comparisons 5, 6, 7 and 9.

It was considered that these criteria, used in both practical and simulation data sets, would give a more powerful and global indication of comparative performance than the single criterion of the mean squared error of the coefficients derived in simulation studies; it is this latter criterion which has dominated most theoretical discussions of the comparison between OLS and ridge.

What follows are the results obtained when two sets of real data and two

sets of simulation data were employed to assess the comparative performance of the two regression methods. A variety of combinations of variables and cases were used to evaluate changes in relative performance. From these the following conclusions can be advanced:

- a. V-ridge is generally robust and flexible and is not usually susceptible to major changes in results with minor changes in the regressor matrix; although error in the model and the nature of the ridge process itself lead to a certain degree of inaccuracy in the V-ridge coefficients, these coefficients are seldom seriously in error.
- b. The smaller the error variance in a model, the better the OLS performance; on the other hand when prediction levels drop to 60-70 per cent or below, V-ridge tends to equal or outperform OLS; at still higher error levels V-ridge performance is nearly always better.
- c. The effect of varying degrees of multicollinearity has not yet been clearly established.
- d. The number of regressors present in a model does not have a clear effect on comparative performance, except in so far as the error structure is affected and thereby the differences in performance.
- e. In general, V-ridge offers a relatively stable level of biased performance within which a reduction in the number of regressors in a model usually causes only minimal changes in coefficient values, while an increase in the error level brings about only a gradual change in these values. In contrast, such alterations in the regressor matrix often result in dramatic changes in OLS coefficient values.
- f. While V-ridge sometimes outperforms OLS by an extremely wide margin, in situations where OLS is superior the OLS advantage is relatively limited.
- g. In a series of runs where V-ridge generally outperforms OLS, the exceptions which occur arise when there is unusually good OLS performance rather than any gross decline in V-ridge performance.
- h. The more the cross-validation samples differ, the better the performance of V-ridge tends to be and the poorer the performance of OLS.
- i. While OLS coefficients are the same whether the regression is based on the covariance or correlation matrix, the results from V-ridge differ to a modest but not critical extent according to the matrix used.
- j. The statistically more important coefficients tend to be estimated more accurately by OLS at the expense of high error in the less important coefficients; V-ridge offers a more balanced set of estimators.
- k. OLS attributes most of the predicted variance to a few important predictors

whereas V-ridge points to a wider spread of unique variance contributions.

1. In view of the ridge 'bias' built into V-ridge coefficients, the most unexpected finding has been that in the Monte Carlo sets evaluated to date the V-ridge coefficients are generally more accurate than those of OLS whenever the level of prediction of the criterion is below 60 or 70 per cent.

The following pages present an outline of the programme development of the Vinod algorithms; a description of the data sets and the results obtained from applying the developed programmes to these sets; and a theoretical discussion of the criteria used for comparing V-ridge and OLS performance.

#### 5.231 Rationale underlying the development of the Vinod algorithms

As already indicated, the development of a working model of ridge regression arose from the need to derive reasonably reliable regression coefficients for the study's research design, where the relative 'weights' or variable contributions to the path analyses would serve as part of the evidence for or against the research hypotheses. A study of most of the available papers published on the ridge method in Britain and the United States suggested to the author that the Vinod model is possibly the only theoretical approach which has an automated non-stochastic algorithm satisfying the demand for a ridge technique that would yield repeatable results when applied by different researchers.

A computer routine for obtaining  $k$  figures for any particular value of  $m$ , using simply the eigenvalues as input data, was kindly supplied to the author by Vinod. This routine makes use of Newton's method of successive approximations to derive  $k$  from  $m$  (the formula cited earlier in this paper shows  $\underline{m}$  as derived from a particular value of  $k$ ).

Early applications of the routine offered one promising indication of performance: a level of coefficient stability, in different sub-samples, well above that of equivalent OLS coefficients. Based on these findings and on other evidence of the statistical potential of the Vinod routine, a year's work was put into developing two extensive computer programmes which incorporate and build on this routine as well as offering a variety of parameters by which the performance of V-ridge can be compared with that of OLS; at the core of the two linked programmes are procedures for three varieties of cross-validation - on odd-even sub-samples, on the top and bottom halves of the data set, and on identifiable sub-groups within the sample. There is also provision for comparing OLS and V-ridge regressions based alternatively on covariance or correlation matrices.

A particular feature built into what is also intended as a research instru-

ment is a procedure for the equal scaling of variables in analyses based on the covariance matrix. As Thisted (1976) points out, there is no reason for adhering, as has been done ever since the original articles were published by Hoerl and Kennard, to the concept that ridge will only operate on a correlation matrix. But Thisted stresses the need to scale variables used in a covariance matrix.

It is clear that without provision for equal or approximately equal scaling, covariance-based ridge could yield bizarre results on a set of variables with widely differing score ranges. The programme's facility for automatic equal scaling also provides the opportunity for differential scaling, if desired, or for basing the equalisation of a variable's scale range not on one or two extreme scores at each end but on a selected range encompassing the majority of scores, as determined from the variable histograms which are also available as output. The selection of a particular scale does not, however, exclude any cases from the data set or analyses. More work is required to determine this facility's potential for V-ridge improvement as well as its theoretical implications.

Provision is also made for disattenuation of the matrix, if required, in terms of available reliability figures on the variable measures. The original and any disattenuated matrices are automatically tested for inversion; if this fails an adjustable small ridge can be added to the diagonal to enable inversion to occur - where for example a particular set of variables may be seen as important despite the ill-conditioning of their matrix. This ridge 'correction' to enable inversion is only an option, however, since it clearly brings about a minor change in the diagonal of the matrix prior to both the V-ridge and OLS solutions.

The major feature of the first of the two computer programmes is the routine to derive Vinod's  $k$ . Linked to this is another essential component, a routine to derive Vinod's Index of Stability of Relative Magnitudes (ISRM) for each value of  $k$ . This is followed by a plot of the log of each ISRM against the corresponding log of  $k$ . The plot provides a characteristic V-shaped dip, with the base of the dip indicating the lowest level of log ISRM. The programme repeats the iterations around this point in two successive series of runs, sharpening the identification of the lowest ISRM to 25 times the original accuracy, thus yielding the 'best' Vinod  $k$  for the main set of data. The entire routine is repeated automatically to derive the best  $k$  figures for the two cross-validation samples.

A comparison of the ordinary least squares and V-ridge regression routines forms the bulk of the second programme. This programme makes use of the  $k$  figures, the stored matrices, data and other parameters derived in the first programme; it yields the OLS and V-ridge regression coefficients for each

data set, together with the residual mean square, the dependent variance accounted for, the standard errors of both sets of coefficients and a considerable number of other results, together with specific comparisons of the performance of OLS and V-ridge regression within each data set and particularly in the cross-validation sets.

An experimental feature built into the second programme is an optional provision for basing the V-ridge regression not on a uniform  $k$  but on individualised  $k_i$  values related to the size of the regressor variances or alternatively based on several other criteria. This feature has yet to be tested extensively; the results of various methods of individualisation await fuller examination. To date the indications are that in one method, that of individual  $k_i$  based on regressor variances, where

$$k_i = \left[ \frac{\text{Mean} (\text{Var} (x_j))}{\text{Var} (x_j)} \right] \cdot k \quad j = 1, 2, \dots, m,$$

the individualisation frequently, though not always, gives a modest improvement above constant  $k$  values in V-ridge.

The second programme offers a fair number of criteria by which the performance of OLS and V-ridge can be compared formally. The difficulty of comparison is that for different theoreticians and analysts there are widely differing criteria by which the performance or superiority of contending methods may be judged.

Reference has already been made to the emphasis placed on mean square error of the regressor coefficients, in most of the ridge studies to date. This characteristic of the estimators has been widely explored by others in simulation studies. The present study concentrates on some of the many other characteristics of the estimators, especially on those parameters of more immediate importance to the analyst.

One criterion which is generally thought to be important is the proportion of total dependent variance accounted for. For reasons related to the nature of the ridge regression algorithm, the V-ridge 'variance accounted for' cannot be determined from the total and residual sums of squares. Accordingly this programme derives the required variance (for both OLS and V-ridge) from the Snedecor and Cochran (1967) formulation:

$$R^2 = \frac{\sum Y_i \hat{Y}_i - ((\sum Y_i)(\sum \hat{Y}_i)/N)}{(\sum (Y_i)^2 - ((\sum Y_i)^2/N))^{\frac{1}{2}} (\sum (\hat{Y}_i)^2 - ((\sum \hat{Y}_i)^2/N))^{\frac{1}{2}}}$$

where  $\sum$  is the summation from  $i = 1$  to  $i = m$   
 $m$  being the number of regressors  
 $N$  is the number of cases

$Y_i$  are the criterion scores  $i = 1, 2, \dots, m$   
 $\hat{Y}_i$  are the predicted scores

All too often this or other derivations of  $R^2$  are cited as the 'variance accounted for', although in fact the estimate has to be corrected for  $N$  and  $m$  in order to yield the variance itself. McNemar (1969) presents the correction as

$$R^{*2} = 1 - ((1 - R^2)(N - 1) / (N - m - 1)) .$$

Any references to variance accounted for are based on this corrected figure.

A second criterion which is also seen to be important is the residual mean square remaining after fitting the model. This is yielded by the usual statistic

$$\frac{\sum (\hat{Y}_i - Y_i)^2}{N - m - 1}$$

A third criterion, one which is used less frequently, is the accuracy with which OLS and ridge regressions can reproduce a set of coefficients in a simulated or Monte Carlo situation. In view of the bias given to the ridge coefficients by the addition of  $k$  to the diagonal, this criterion should be of particular interest. Clearly an inability to approach the true coefficients in a simulation data set should be seen as a particular disadvantage when examining a set of estimators, however well the total variance may be accounted for or whatever the minimisation of the residual mean square (RMS). A relatively simple statistic has been used here to define the coefficient inaccuracy:

$$\text{Inaccuracy} = \frac{\sum |((b_{iT} - b_{i0}) / (b_{iT}))|}{m}$$

where  $b_{iT}$  is the 'true value' used in the generation programme  
 $b_{i0}$  is the derived value of the coefficient

An inaccuracy statistic greater than unity is clearly an indication of serious error in the derived coefficient set, seen as a whole.

A fourth and perhaps the most important measure of all has not been used in evaluating ridge estimators up to now and appears only rarely in other forms of model or method assessment. Stone (1974) and a number of earlier writers cited in his study emphasise the necessity of cross-validation, especially when assessing new techniques. In one form of this procedure data are randomly or deliberately divided and the coefficients or other statistics derived from one data set are tested on the other data set.

Within the cross-validation process as developed for this study there are



five subsidiary criteria of performance which merit particular attention:

a. The mean overall predictive power of the V-ridge constant and coefficients, compared to those of OLS, when the parameters derived from each sub-sample are applied to the data in the other sub-sample. The criterion is based on the variances accounted for in the cross-prediction situation.

b. The mean percentage decrease (or increase) in RMS resulting from V-ridge, as compared to OLS in the cross-validation. This is based on:

$$(RMS_{V\text{-ridge}} - RMS_{OLS}) / (RMS_{OLS})$$

c. The separate coefficient stability levels for V-ridge and OLS coefficients; these are based on the statistic

$$\frac{\sum \left\{ |(B_{1i} + B_{2i}) / 2| - \left| ((B_{1i} + B_{2i}) / 2) - B_{1i} \right| \right\}}{\sum |(B_{1i} + B_{2i}) / 2|}$$

where  $B_{1i}$  and  $B_{2i}$  are the standardised coefficient values (for regressions based on correlation matrix) or deviation values (for regressions based on covariance matrix), derived in the first and second cross-validation samples.

(Earlier formulations of this statistic were simpler but tended to yield extreme values, especially when the mean sizes of one or more coefficients were close to zero. The 'averaging' effect achieved above provides a more balanced estimation of the OLS and V-ridge stabilities.)

d. The overall or total stability across both sets of V-ridge coefficients compared to that across both sets of OLS coefficients; in this case the size of the total differences across pairs of coefficients in the two V-ridge sets is compared with that across the two OLS sets, in the form

$$\frac{\sum (\text{abs. difference between pairs of OLS } B_{1i}, B_{2i} \text{ across sub-samples})}{\sum (\text{abs. difference between pairs of V-ridge } B_{1i}, B_{2i} \text{ across s.samples})}$$

In effect the statistic is equivalent to the overall stability of V-ridge divided by that of OLS, i.e. the inverse of the relative instability as defined here. In a situation where OLS and V-ridge achieve approximately the same levels of prediction and RMS, the more parsimonious set of coefficients may be thought to be more effective. Differences between pairs of coefficients bring out this feature.

e. A combined statistic, termed the Index of V-Ridge Effectiveness (IVE). This index is simply the empirical product of the four previous cross-validation indicators. In effect it has the form

(Predictive Effy.) X (RMS change) X (Separate stability) X (Total stab.)

A further cross-validation criterion, the maximum and minimum single coefficient discrepancies across the subsets, is determined separately for OLS and V-ridge, but is not included in the combined IVE.

The relative ability of a set of regression coefficients to maximise overall predictive power and to minimise residual mean square should clearly be the main criteria in judging the effectiveness of a regression method in those research situations where the aim is to link a particular regressor matrix as closely as possible to its criterion variable or to predict individual scores in a clinical trial, for example. If on the other hand the aim is to set up a general model or to discover variable parameters which have a high degree of validity when applied to other sample sets or to sub-samples within the original data set, then it is the accuracy of the coefficients and the cross-validation stability measures which may yield more useful evidence about the effectiveness of differing methods of regression.

In addition to the cross-validation measures the programme yields t-statistics based on the standard errors of the regression coefficients derived from the main data set. For the OLS coefficients the statistic is obtained in the ordinary way as

$$t_j = b_j / ((\text{RMS})(C_{jj}))^{\frac{1}{2}},$$

where  $C_{jj}$  is the diagonal  $jj$  of the inverse regressor matrix.

For the V-ridge coefficients it was necessary to use a more basic formulation:

$$t_j = b_j / (\text{diag. } jj \text{ of } ((\text{MSE})(\text{Var}(B)))^{\frac{1}{2}},$$

where

$$\text{Var}(B) = (X^t X + kI)^{-1} X^t X (X^t X + kI)^{-1}.$$

Another criterion, the unique variance predicted by each variable and the remaining shared variance predicted in common, is also derived in the programme. Details are not given here.

## 5.232 Results

The results of the comparisons between OLS and V-ridge on four different data sets are presented in this study. Two of these are simulation sets, the data having been generated in a programme written by Chalmers (1979); this programme was based on a random number input into a generator routine for which the regression coefficients, the approximate correlation matrix, the standard deviations and the random error structures were specified. The two other sets comprise the following: a real data set obtained by courtesy of Gary McDonald, of General Motors Research Laboratories, Warren, Mich., U.S.A., this set being the same as that used in a regression analysis of mortality and pollution data (McDonald and Schwing, 1973); and a real data set taken from research recently completed by the present author.

A. Simulated intervention data. The parallel sets of OLS and V-ridge coefficients derived in one of the runs on 159 cases in the main sample employed for the present study were averaged to give a mean coefficient mid-way between the values yielded by the two methods. The regressors consisted of seven educational, environmental and psychological variables. The correlation matrix (not disattenuated in this case) was taken from the same sample. The mean coefficient values and correlation matrix (reproduced in table 8) were used as input into the Chalmers programme to generate 200 cases with varying levels of error structure. The medium error used for this first analysis was approximately equal to the error levels found in the original data. In view of the random methods used in generating this and subsequent data sets, sub-samples were drawn by simple procedures such as taking every second case if, for example, a sample of 100 cases was required.

Table 9 presents a comparison of OLS and V-ridge results on the simulation, the variant here being the number of cases, ranging from the full data set of 200 cases to a subset of 34 cases; in the latter sample the cross-validation sub-samples consisted of 17 cases each. Eight of the comparison criteria feature in this table.

The ninth criterion is a comparison of the t-statistics for the sets of derived coefficients. The two sets used in table 10 (p. 297) are those where the OLS coefficients have their lowest and their highest inaccuracy statistics; the comparable V-ridge coefficients are taken from the same sets. The t-statistics appear in brackets below the coefficients.

LEGEND for tables 8 to 17

The uniform set of symbols and criteria used in the nine tables in this paper are as follows:

OLS: Ordinary least squares regression parameters or estimators

V-R: V-ridge regression parameters or estimators

V/O: Ratio of V-ridge to OLS estimators

RMS: Residual mean square of predicted criterion scores

% RMS change: Per cent change in RMS for V-ridge (relative to OLS)

VAF: Variance accounted for by derived regression coefficients

Inacc: Inaccuracy in replicating simulation coefficients

Main RMS, Main VAF: RMS and VAF for main sample

Subsets: defined as odd/even; or top/bottom halves; or specified groups of the main sample

Coefft stability: separate and total stabilities of regression coefficients, stabilities being determined as described in text

IVE Index of V-Ridge Effectiveness; a composite measure summarising the VAF ratio, the RMS ratio and the relative stabilities (see text)

\*\* Comparison of V-ridge VAF with that of OLS is normally based on  $R^2$  corrected for N and the number of regressors; when this adjustment reduces either VAF to a negative figure, the particular comparison is based on non-adjusted  $R^2$ .

Ind(v.) V-ridge determination based on individualised k figures

k = when a k figure is noted, the regression routine uses specific ridge k values rather than the automatic k values derived in the Vinod routine

Table 8. Correlation matrix and coefficients for simulation sample set A

	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_7$	
$X_1$	.7436	.8045	.3315	.0150	.4047	.7066	
$X_2$		.7272	.5038	.0456	.3335	.7122	
$X_3$			.2980	-.0073	.3269	.6097	
$X_4$				.2560	.1394	.4674	
$X_5$					-.1138	.1150	
$X_6$						.2249	
Constant	$b_1$	$b_2$	$b_3$	$b_4$	$b_5$	$b_6$	$b_7$
0.0	.124	.174	.215	.051	.037	.024	.049

Standard deviation of regressor variables: 1.0 throughout

Identity of variables:

- |                              |                               |
|------------------------------|-------------------------------|
| 1. Nursery ability cluster   | 2. Nursery attainment cluster |
| 3. Reception ability cluster | 4. Home variable cluster      |
| 5. Programme variable        | 6. Maslow needs state         |
| 7. English Pic. Vocab. Test  |                               |

Criterion (theoretical): Composite reading level

Table 9. Comparison of OLS and V-ridge regression on sample set A (7 regressor variables)

Sample set	Main RMS OLS V-R		Main VAF OLS V-R		Inaccuracy in reproducing coefficients						Cross-validation - mean values				IVE V/O	
					Main set OLS V-R		Subset A OLS V-R		Subset B OLS V-R		VAF V/O	% RMS change	Coefft. stability			
													Separate OLS V-R	Total V/O		
N 200	.17	.18	.64	.63	.86	.49	.99	.65	.89	.44	.98	+5.5	.75	.88	2.50	2.76
N 150	.18	.19	.63	.61	.80	.51	.52	.58	1.62	.62	1.04	-5.9	.43	.76	2.93	5.66
N 100	.18	.20	.61	.60	.99	.62	1.37	.84	.92	.49	1.03	-3.2	.59	.76	2.29	3.12
N 67	.16	.17	.66	.64	1.14	.49	1.92	.51	1.60	.76	1.18	-16.1	.20	.68	3.94	19.25
N 50	.19	.20	.49	.47	1.37	.87	1.68	.95	2.16	1.18	1.56	-18.5	-.19	.48	3.06	***
N * 34	.10	.13	.75	.68	1.82	.57	2.09	.81	2.19	.59	.94	+5.7	.51	.56	2.07	2.03

Subsets based on odd/even division of main sample

\* Comments on this set of statistics appear elsewhere

\*\*\* Impossible to calculate since OLS stability level was below zero (deviations exceeded mean coefficient values)

Table 10. Comparison of coefficient stability across OLS and V-ridge solutions

'True' constant and coefficients (in generation programme)							
0.000	.124	.174	.215	.051	.037	.024	.049
OLS constant and coefficients for N = 150							
0.004	.038 (0.54)	.115 (1.83)	.343 (5.40)	-.021 (0.47)	.037 (1.11)	.079 (1.98)	.035 (0.60)
V-ridge constant and coefficients for N = 150							
0.001	.087 (4.98)	.095 (5.22)	.212 (8.93)	.019 (0.52)	.023 (1.13)	.062 (1.98)	.053 (0.60)
OLS constant and coefficients for N = 50							
0.063	.006 (0.04)	.079 (0.070)	.332 (2.48)	-.023 (0.30)	.045 (0.75)	.129 (1.91)	-.025 (0.22)
V-ridge constant and coefficients for N = 50							
0.060	.058 (1.77)	.076 (2.38)	.175 (4.07)	-.001 (0.04)	.028 (0.71)	.095 (2.26)	.021 (0.61)

While the criterion statistics in table 10 must in general speak for themselves, a few comments could be added. Both the inaccuracy statistics and the separate coefficient stability figures show that there is a reasonable consistency in the V-ridge results compared with the very considerable fluctuations in the OLS results. This feature is the most pronounced characteristic of V-ridge - the relative steadiness of its estimators, with extreme error being found only rarely. The first four indicators of RMS and VAF (variance accounted for) on the main data set suggest why many writers choose to reject the credentials of ridge regression and in particular that of Vinod's ridge  $k$ ; since V-ridge is not based on manipulation of the ridge trace to obtain the 'best' mean square error on the original data base, it is likely to do less well than any stochastic ridge method when criteria of performance on the original data set are the only ones considered.

The cross-validation runs show that the main strength of V-ridge lies in its coefficient stability across the two sub-sets, and this strength is born out by the major advantage of V-ridge in the accuracy of reproduction of the simulation coefficients. (Later examples will show how the superiority or inferiority of V-ridge on the combined Index of V-Ridge Effectiveness (IVE) is closely related to the relative levels of V-ridge and OLS inaccuracy.) Even in cases where the OLS values for RMS and VAF in the cross-validation are better than those of V-ridge, the V-ridge inaccuracy figures are between one-third and one-half

smaller than those of the OLS figures.

The final set of results cited in table 9 is an interesting one and reflects the complexity of the statistical features of regression parameters. The OLS cross-validation performance is superior to that of V-ridge on two out of the four indicators cited - RMS and VAF. The main data set's RMS and VAF figures are also quite an improvement on the V-ridge figures. Yet the inaccuracy statistics show that the OLS coefficients in the main and in both the subsets are very far from the 'true' values. It could be argued that this is due to the structure of the remaining 34 cases having moved a great distance from the structure of the original sample of 200 cases, which would be a reasonable possibility. This would not account for the fact that V-ridge on the same 34 cases has succeeded in offering a reasonably close estimate of the 'true' coefficients (inaccuracy figures of .57, .81 and .59 compared to the OLS figures of 1.82, 2.09 and 2.19). What appears to have occurred here and in some other situations which have been examined is that the ordinary least squares algorithm succeeds in capitalising on certain features in the regressor matrix to produce high VAF and low RMS statistics, even in subsets drawn from the main data set, and furthermore produces coefficients which are (in this case) reasonably stable across the two subsets, and yet the coefficients themselves are very far from the true values.

B. Real intervention data. A real data set was taken from the present research sample of 159 cases. On this occasion six variables were used as predictors, with English Picture Vocabulary Test as the criterion. The purpose of this series was to indicate briefly the effects of different types of cross-validation. Only the results are published here, in table 11.

The comparison between OLS and V-ridge regression based on either covariance or correlation matrices suggests that for this series the covariance based V-ridge performs better than does the correlational equivalent. A more important finding, in theoretical terms, is that when the data subsets are formed from somewhat differing groups, rather than from odd/even or top/bottom halves of the main data set, comparative V-ridge performance appears to be considerably improved. The sample groups were formed on the basis of programme attendance or non-attendance; it could be expected that the characteristics of parent and child performance would differ fairly considerably between the 99 programme attenders and the 60 non-attenders. In the cross-validation the coefficients derived by V-ridge differed across the groups but still bore a reasonable resemblance to each other - again as could be expected. The depressed performance of OLS in this situation, as indicated by the IVE performance figures, suggests again



Table 11. Comparison of OLS and V-ridge regression on sample set B (6 regressors)

Sample N      Sub-sets		Main RMS OLS    V-R		Main VAF OLS    V-R		Cross-validation - mean values					IVE V/0
						VAF V/0	% RMS change	Coefft. stability			
								Separate OLS    V-R	Total V/0		
159*	Odd/ even	35.7	40.9	.61	.60	1.03	+5.5	.69	.84	2.54	3.00
159*	Groups	35.7	40.9	.61	.60	1.06	-3.1	.46	.74	2.70	4.73
159*	Top $\frac{1}{2}$ / Bottom	35.7	40.9	.61	.60	1.04	+6.8	.67	.83	2.62	3.15
159	Odd/ even	35.7	41.6	.61	.59	0.99	+8.5	.73	.84	2.32	2.46
159	Groups	35.7	41.6	.61	.59	1.03	-1.7	.49	.74	2.54	3.97
40*	Odd/ even	51.1	57.3	.41	.39	1.48**	-26.5	.20	.65	3.76	24.29
40*	Groups	51.1	57.3	.41	.39	2.26**	-40.4	.13	.45	2.35	31.21
40*	Top $\frac{1}{2}$ / Bottom	51.1	57.3	.41	.39	0.90	+8.2	.84	.76	1.12	0.85
40	Odd/ even	51.1	57.2	.41	.39	1.44**	-25.3	.15	.61	3.74	29.98
40	Groups	51.1	57.2	.41	.39	1.77**	-35.6	.11	.43	2.21	22.89
40***	Odd/ even	51.1	57.2	.41	.39	1.50**	-26.5	.20	.67	4.14	28.02

\* These regressions based on covariance matrices; non-starred runs were based on correlation matrices

\*\* Based on uncorrected  $R^2$  ( $R^2$  between predicted and actual values of criterion variable) rather than on variance accounted for

\*\*\* In this covariance-based run the V-ridge regressions were based on individualised k values; results may be compared with the first run on 40 cases

that the OLS regressions are emphasising features in the regressor matrices that are largely irrelevant to the accuracy of the coefficients. It appears that the weighting of the diagonal variances in the V-ridge method succeeds in limiting the least squares capitalisation tendency, though without foregoing the advantages of the basic error minimisation process by which the coefficients are derived.

The accuracy of V-ridge coefficients when derived from differing groups within a sample has been noted throughout the study. Its importance is that in most research situations the coefficients that are derived from a particular sample are likely to be validated on other samples which are inevitably rather different from the original sample, even if the difference is only one of a secular trend covering two successive annual samples, for example. It is less often that coefficients are required for use on random samples drawn from exactly the same population that yielded the original sample. The further apart the samples are the stronger the advantage of V-ridge is likely to be, provided the data sets are small to medium in size. The comparable situation in large data sets is still under investigation.

C. Simulated mortality data. The origin of this data lies in a group of 16 social and environmental variables which were measured in 60 different areas of the United States. They included a number of air pollution variables. The mean of each variable was determined for each area and these figures were then used by McDonald and Schwing (1973) in a regression equation predicting mortality levels. The authors compared the performance of OLS and ordinary ridge methods (using an examination of the ridge trace in order to decide on  $k$  values) and came to the conclusion that the two methods each pointed to six important predictors, although the methods agreed only on five of these predictors; thus in total there were seven variables which either OLS or ridge identified as important according to the criteria used in their study.

The relevant OLS and ridge coefficients derived in that study were used to generate a set of 200 cases for the present study. For the five common predictors the means of the OLS and ridge coefficients were used; for the other two predictors one coefficient was taken from the OLS regression (variable 3, which only OLS showed to be important), and the other from the ridge regression (variable 8, which only ridge showed to be important). It should be noted that since the McDonald and Schwing ridge  $k$  was only 0.2 compared to the V-ridge (correlation matrix) figures of between 0.4 and 0.8, this choice of 'mean' coefficients and the accompanying matrix clearly biases the structure of the simulation data set in the direction of an OLS rather than a V-ridge solution. Table 12 presents the McDonald and Schwing correlation matrix on which the Chalmers generation programme was based, together with the coefficients and standard

Table 12. Correlation matrix and coefficients for simulation sample set C

	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_7$	
$X_1$	.0922	.5033	-.4904	-.0035	.4132	-.1069	
$X_2$		.3463	.1163	-.1001	.4538	-.1078	
$X_3$			-.2385	-.0610	.5753	-.0993	
$X_4$				-.2439	-.2088	-.2343	
$X_5$					-.0057	.4321	
$X_6$						.1593	
Constant	$b_1$	$b_2$	$b_3$	$b_4$	$b_5$	$b_6$	$b_7$
1084.4	1.64	-1.56	-2.36	-12.58	0.004	4.37	0.253
Standard deviations of regressors:							
	9.98	10.17	4.76	0.85	1454.10	8.92	63.39

Identity of variables:

- |                            |  |
|----------------------------|--|
| 1. Mean annual rainfall    | 2. Mean January temperature            |
| 3. Mean July temperature   | 3. Mean educational level of area pop. |
| 5. Area population density | 6. Percentage Non-White in area        |
| 7. Sulphur dioxide level   |  |

Criterion (theoretical): Mortality level

deviations employed in the simulation routine. There were five levels of error built into the five data sets; details of these are given at the bottom of table 13.

Table 13 presents the results of a series of OLS and V-ridge comparisons in which the error levels are systematically varied across samples of two sizes, 134 and 67. In contrast with the previous set of data, the OLS performance here is comparable with that of V-ridge in the statistics on inaccuracy in reproducing coefficients. Where there is low error and consequent high prediction (VAF of .90 and above) OLS performs well above V-ridge both in the inaccuracy statistics and in the cross-validation; the accuracy of OLS in this cross-validation is such that by comparison there is a massive increase in the RMS for V-ridge relative to OLS, although the V-ridge predictive efficiency does not fall below .91 of the OLS figure. At the other extreme of high error and low prediction (.27 to .39) V-ridge outperforms OLS on all the indicators. An interesting finding in table 13 is that the V-ridge inaccuracy statistic increases only moderately with increase in error level - generally less than 50 per cent from the lowest to the highest levels of error in the model. In contrast the OLS inaccuracy statistic grows from a minimal size at low error to figures many hundred per cent greater at the top error levels.

There are indications in the inaccuracy statistics that the part played by variable 3 (July temperature) is of special importance. Although this variable correlates positively with the mortality criterion (in both the simulated and actual data), it yields a negative coefficient. It can arguably be seen as a suppressor variable, whose characteristics have been described by Cohen and Cohen (1975), among others. However, the main ridge trace presented by McDonald and Schwing (ibid) shows that this coefficient reaches zero at high  $k$  levels (the only one of the 15 predictor coefficients to do so). Thus July temperature can reasonably be seen as a variable whose apparently sizable coefficient value is related only to its collinearity with other regressors. The fact that V-ridge consistently predicts values close to zero for this coefficient contributes a fair amount to the relatively poor V-ridge results in the inaccuracy statistics of table 13.

A fuller comparison of the effects of different variable sets was carried out on 100 cases at the medium error level. The results appear in table 14. A random numbers table was used to eliminate one or more of the seven variables entered into each of the 21 regressions subsequent to the regression on all seven variables. The effect of breaking the original structure in which variable 3 fills its main role as a suppressor variable is that in only four of the total of 22 regressions is the OLS inaccuracy statistic lower (better) than that of V-ridge. In the cross-validation series the V-ridge performance is better in 13 of the 22 prediction comparisons, in only 6 of the residual mean

Table 13. Effect of error level: OLS vs. V-ridge in sample C (7 regressors)

s.dev. of $\hat{Y}$ error	Inaccuracy in reproducing coeffs.						Cross-validation values***				IVE  V/0
	Main set		Subset A		Subset B		VAF V/0	% RMS change	Coefficient stability		
	OLS	V-R	OLS	V-R	OLS	V-R			OLS	V-R	
<u>N = 134</u>											
10	.10	.46	.17	.46	.09	.45	0.91	+504.7	.94	.90	0.17
20	.20	.46	.33	.49	.18	.47	0.91	+117.9	.87	.89	0.93
40	.40	.53	.66	.55	.37	.52	0.93	+23.5	.73	.82	2.53
50	.49	.57	.82	.59	.45	.55	0.94	+12.5	.65	.78	3.14
70	.70	.64	1.15	.61	.64	.60	0.98	+3.2	.54	.69	4.05
<u>N = 67</u>											
10	.17	.46	.33	.50	.19	.57	0.89	+409.1	.87	.85	0.28
20	.33	.50	.67	.48	.38	.62	0.91	+82.1	.75	.84	1.95
40	.66	.56	1.34	.54	.76	.72	1.02	+3.3	.54	.72	5.14
50	.82	.59	1.67	.61	.95	.77	1.14	-5.7	.45	.65	6.92
70	1.15	.66	2.34	.74	1.33	.87	1.20**	-13.2	.34	.52	8.53

\* The error standard deviations in the simulation were those used in the Chalmers generation programme; the error levels gave rise to predictive variances in the following ranges (over both main data sets):

10: .90 to .97      20: .80 to .88      40: .53 to .62

50: .41 to .50      70: .23 to .31

\*\* Based on uncorrected  $R^2$  ( $R^2$  between predicted and actual values of criterion variable) rather than on variance accounted for

\*\*\* The total stability figure omitted from this table, but the IRE includes total stability as well as the other three cross-validation comparisons

Table 14. Effect of variable sets: OLS vs. V-ridge in sample C ( N = 100 )

Variables in model	Main RMS OLS V-R		Coefficient inaccuracy (mean of 3)* OLS V-R		Cross-validation - mean values					IVE V/0
					VAF V/0	% RMS change	Coefft. stability		Total V/0	
							Separate OLS V-R			
1 2 3 4 5 6 7	1824	2418	.46	.54	0.91	+15.9	.82	.72	1.24	0.86
1 3 4 5 6 7	2155	2563	.54	.50	1.00	+6.0	.82	.77	1.45	1.29
1 2 3 4 6 7	1813	2362	.30	.53	0.95	+14.1	.82	.75	1.36	1.03
1 2 3 4 5 7	2889	3095	.88	.63	1.15	-3.2	.42	.65	2.31	4.27
1 4 5 6 7	2239	2538	.55	.37	0.99	+3.4	.79	.78	1.36	1.30
1 3 5 6 7	2312	2703	.36	.50	1.00	+6.0	.81	.73	1.15	0.99
1 2 3 4 7	2859	3057	.86	.67	1.10	+0.2	.63	.71	1.86	2.32
1 2 3 4 6	1942	2774	.45	.56	0.89	+20.1	.78	.42	1.16	0.46
2 3 4 5 7	3165	3357	1.06	.69	1.07	+0.0	.62	.67	1.71	2.00
1 2 4 5 7	2918	3139	.75	.41	1.07	-2.2	.46	.61	1.91	2.75
1 5 6 7	2340	2681	.39	.32	0.99	+5.6	.78	.72	1.10	0.95
1 2 3 4	3540	3697	.88	.72	1.41	-5.7	.25	.62	3.02	11.35
2 4 5 7	3316	3567	.92	.38	1.03	+1.5	.55	.67	2.11	2.59
1 5 6	2762	3079	.48	.44	0.97	-1.2	.74	.73	1.45	1.41
1 5 7	2976	3329	.88	.26	1.00	+4.3	.61	.69	2.00	2.16
1 2 3	3783	3956	1.15	.96	1.52	-3.6	.39	.59	3.35	7.89
1 3 4	3527	3706	.85	.64	1.03	0.5	.65	.75	2.03	2.38
2 4 5	3784	3972	1.04	.64	1.11	-2.3	.39	.56	2.25	3.70
4 5 7	3288	3547	.84	.15	0.98	+5.8	.74	.80	2.04	2.05
1 5	3614	3870	1.21	.50	0.99	+3.5	.65	.68	1.87	1.87
1 3	3795	3985	.77	.27	0.97	+2.5	.67	.70	1.78	1.76
4 5	3745	3942	.99	.75	0.99	+2.5	.68	.79	2.45	2.73

\* Mean of main set and two subsets (odd and even halves)

- Notes: a. Error level was 40 (see table 5); predicted variance  
b. Regressions based on covariance matrix  
c. Individualised k (variance-based) used in V-ridge regressions

Variables: 1. Mean ann. rainfall; 2. Mean Jan. temp; 3. Mean July temp;  
4. Mean educa. level; 5. Pop. density; 6. Percent Non-White;  
7. Sulphur dioxide level. Criterion (theor.): mortality level.

square comparisons, in 14 of the comparisons based on separate stability figures and in all 22 of the total stability measures. Thus even in a data set where ordinary least squares capitalises on the performance of a suppressor variable, V-ridge outperforms OLS in the majority of the situations where the unique conditions of the original data set no longer pertain.

The question remains however as to whether variable 3 is a genuine suppressor variable whose contribution to the model is only recognised by OLS. It is a problem that merits further consideration.

D. Real mortality data. The original data set used by McDonald and Schwing (ibid) was supplied to the author by McDonald (personal communication). The data appear in McDonald and Ayers (1974). These data formed the basis of a wide range of regression runs in the present study, using different variable sets; some of these were the same as the variable sets cited in the McDonald and Schwing study. The stochastic k figure of 0.2 which the authors considered to be the most appropriate value for ridge regression was also used in some of the runs, to compare performance with that of V-ridge where the k figures were invariably much higher.

Table 15 presents the means, standard deviations and correlation coefficients from these data (60 cases). Clearly they are the same as those derived in the McDonald study. The original computer programme for V-ridge only provided for a maximum of 11 regressors and thus the initial runs were carried out using variables 1 to 11 and variables 5 to 15; this indicates fairly clearly, in agreement with McDonald's findings, which variables are worthy of further study.

Tables 16 and 17 present the results of the comparisons between OLS and V-ridge. The performance of V-ridge is far above that of OLS in the first five runs in table 16. This would not necessarily occur with every set of 11 regressors, but it does indicate the seriousness of the over-capitalisation by ordinary least squares on chance features of the matrix. The choice of the 'best 11' and then the 'best 8' regressors yields more conventional results.

Various forms of cross-validation samples are used in these runs. The cross-validation in the first five runs is based on the top and bottom halves of the 60 cases. Three runs are based on what were hypothesised to be somewhat different groups within the sample - communities divided according to their proportions of white-collar populations, mortality rates and years of education completed. While V-ridge maintains a consistent level of performance in all three runs, in one of the runs (white-collar groupings) the OLS performance is unusually good and supersedes that of V-ridge.

**Table 15. Basic parameters of sample set D (60 cases)**

[illegible]



Table 16. Comparison of OLS and V-ridge regression on sample set D (60 cases)

Selection of regressors	Regression model	Main sample		Cross-validation - mean values					IVE  V/0
		VAF V/0	% RMS change	VAF V/0	% RMS change	Coefft. stability			
						Separate OLS	V-R	Total V/0	
1 to 11 (11 regr.)	Corr; Ind. Top/bottom	0.88	+35.8	1.31**	-50.8	.30	.62	4.82	26.12
1 to 11	Corr; Top/bottom	0.87	+37.5	1.24**	-47.2	.30	.62	4.65	22.03
1 to 11	Covar; Ind. Top/bottom	0.89	+35.8	1.39**	-53.4	.30	.60	4.89	29.04
1 to 11	Covar; Top/bottom	0.87	+36.8	1.31**	-50.8	.30	.59	4.50	23.59
5 to 15 (11 regr.)	Covar: Ind.	0.87	+32.6	2.08**	-62.7	.29	.51	8.32	80.20
Best 11	Covar; Ind.	0.90	+37.0	1.22**	-37.9	.79	.59	3.18	4.70
Best 8	Covar; Ind.	0.91	+40.7	1.35	-16.1	.51	.64	2.74	5.49
W.col.gps (best 8)	Covar; Ind. Groups	0.91	+40.7	0.57	+56.6	.75	.49	1.00	0.24
Morty.gps (best 8)	Covar; Ind. Groups	0.91	+40.7	0.93**	+77.9	.33	.41	3.91	2.53
Educa.gps (best 8)	Covar; Ind. Groups	0.91	+40.7	6.44**	-51.1	.37	.50	2.43	42.84
McDon. A (best 7)	Corr; Ind. V-ridge k	0.94	+30.4	0.93	-3.8	.64	.63	1.57	1.49
McDon. A	Corr; V-ridge k	0.92	+49.5	0.84	+6.8	.64	.65	2.09	1.67
McDon. A	Corr; k = 1.0	0.91	+57.6	0.81	+12.0	.64	.64	2.28	1.66
McDon. A	Corr; k = 0.6	0.94	+35.2	0.87	+2.6	.64	.64	1.85	1.56
McDon. A	Corr; k = 0.4	0.96	+22.5	0.91	-2.5	.64	.64	1.61	1.50
McDon. A	Corr; k = 0.2	0.98	+9.2	0.96	-6.6	.64	.64	1.34	1.38
McDon. A	Corr; k = 0.1	0.99	+3.2	0.99	-6.5	.64	.64	1.18	1.25

\*\* Based on uncorrected  $R^2$  ( $R^2$  between predicted and actual values of criterion variable) rather than on variance accounted for

Cross-validation: when groups specified, main sample divided into White collar / non-W.C.; high/low mortality; high/low years education

top/bottom indicates division into top and bottom halves

when neither specified, divided into odd/even halves

Best 11 and best 8: regressors contributing most to prediction (seeking optimum agreement between OLS and V-ridge) out of total of 15 regrs.

McDon. A: 'best 7' variables selected by McDonald and Schwing (see text)

Table 16 (continued): Comparison OLS and V-ridge - sample set D (60 cases)

Selection of regressors	Regression model	Main sample		Cross-validation - mean values					IVE  V/0
		VAF V/0	% RMS change	VAF V/0	% RMS change	Coefft. stability			
						Separate OLS	V-R	Total V/0	
McDon. B (6 regr.)	Covar;	0.93	+44.4	0.89	+14.5	.71	.70	1.99	1.53
McDon. B	Corr; Ind.	0.94	+31.3	0.93	+8.0	.54	.68	2.29	2.48
McDon. B	Corr;	0.93	+47.0	0.92	+15.1	.54	.67	2.66	2.68
McDon. B	Corr; k = 0.2	0.98	+8.8	0.99	-2.9	.54	.66	1.68	2.12
McDon. C (6 regr.)	Covar;	0.95	+33.6	0.94	+8.0	.66	.72	2.07	1.95
McDon. C	Corr; Ind.	0.97	+20.11	0.97	-1.4	.63	.65	1.46	1.49
McDon. C	Corr;	0.96	+36.4	0.89	+10.0	.63	.70	2.06	1.86
McDon. C	Corr; k = 0.2	0.99	+5.9	0.97	-2.9	.63	.65	1.30	1.35
McDon. D (5 regr.)	Covar;	0.96	+31.8	0.94	+16.0	.71	.74	1.89	1.60
McDon. D	Corr;	0.97	+34.0	0.97	+16.4	.65	.73	2.11	1.97
McDon. D	Covar; k = 0.2	1.00	0.0	1.00	-0.1	.71	.71	1.003	1.01
Mean temps (2 regr.)	Covar;	0.94	+3.4	1.04	+0.2	.65	.79	3.69	4.61

Cross-validation samples: all based on odd/even division

k values: where k is unspecified, the automatic V-ridge k is used

McDon. B, C and D: 'best' variable sets, based on McDonald and Schwing selections (see text)

Particular attention is paid in these runs to the variables identified by McDonald and Schwing as important. The runs labelled 'McDon. A' are based on the best OLS and ridge sets, yielding 7 predictors in all. (These are the seven variables used in the previous simulation exercise.) Comparisons are made here not only between OLS and V-ridge but also between OLS and ordinary ridge  $k$  figures, the latter being varied between 1.0 and 0.1 and including the figure of 0.2 which those authors considered to be the preferred  $k$  on the basis of the ridge trace.

'McDon. B' is the set of runs based on the best OLS regression, according to their criterion of 'standardised total squared error', a function of the residual sum of squares from each fitting equation; six predictors were identified in this way. A slightly different set of six variables (variable 8 instead of variable 3) were identified by the authors' ridge regression procedure and based on the same criteria as above; these variables are used in the runs labelled 'McDon. C'. The final set of runs, 'McDon. D', are based on the five variables identified as important by both the OLS and ridge regressions.

In each of these four sets of variables the runs based on  $k = 0.2$  yield a less satisfactory performance, in terms of the overall Index of V-Ridge Effectiveness (IVE), than the runs based on the automatic V-ridge ( $k$  being generally in the neighbourhood of .65 to .85 for these runs). An examination of the conventional (Hoerl and Kennard) ridge trace presented in the McDonald study indicates that although some stabilisation of coefficient values occurs between  $k = 0$  and  $k = 0.2$ , the last reversal of coefficient sign occurs just before  $k = 0.6$ . The Vinod  $k$  figures are evidently at a high enough level to overcome this 'delayed correction' of one of the coefficient signs.

As a matter of interest, a single run was carried out using only the January and July temperatures as predictors of mortality. The V-ridge performance remains considerably above that of OLS, although the total prediction is relatively low, as could be expected.

The final runs on this data were carried out on all 16 variables, using an enlarged V-ridge computer programme. (This version can handle the dependent variable and 15 regressors - possibly the maximum that could reasonably be entered into a regression before one is forced to group regressors into composite variables.) A regression based on the covariance matrix yields an IRE statistic of 28.89. This unusually high figure is made up of a V-ridge predictive efficiency of 1.46 compared with OLS, a V-ridge decrease of 44 per cent in RMS, a separate stability level of .57 for V-ridge compared to .48 for OLS and a total stability ratio of 9.20. The maximum and minimum deviations of paired coefficients in the cross-validation also suggest the overall stability here of V-ridge, with a maximum deviation of 3.15 for OLS compared to 0.21 for V-ridge

and a minimum figure of 0.0488 for OLS compared to 0.0057 for V-ridge. The comparative regressions based on the correlation matrix yields an even higher IRE figure of 35.96; one of the routines for the individualisation of k raises this figure to 38.20.

In normal circumstances the results from the regression on the 15 potential predictors would be used to decide on successive small reductions in the number of predictors, carrying out a new regression after each reduction until the final acceptable set has been reached. There are various methods of deciding on these reductions. The simplest is to rely on the set of t-coefficients yielded by each regression. These have been used here to compare the results from the 15-variable regression with those from the McDon. B and McDon. C sets of regressors (the sub-sets most favoured by V-ridge and OLS respectively).

In the 15-regressor run five V-ridge coefficients yield significant t-figures (precipitation, January temperature, years of education, per cent Non-White and sulphur dioxide potential), with population density yielding a near-significant t of 1.95 (p .05, 2-tailed, N=60). The V-ridge regressions on the reduced sets of 6 regressors in the McDon. B and McDon. C runs give coefficient significance figures which are reasonably close to those found in the original set of 15 regressors, although the t-statistic for January temperature is reduced from -2.21 to -1.80 in both the smaller regressions. There are however no major changes either in the significance figures or in the coefficients themselves.

In contrast the OLS regression on 15 variables yields only two significant t's (precipitation and per cent Non-White), with the t of -1.74 for January temperature approaching significance. On the two McDonald subsets the OLS regressions provide five significant coefficients and one of near significance. For example the t of sulphur dioxide changes from 0.57 in the 15-regressor run to 2.83 and 3.28 in the subset runs.

Another feature which distinguishes the two comparisons is that there are no highly valued V-ridge coefficients in the 15-regressor run which are later shown to be meaningless, whereas the comparable OLS run attributes relative importance to three variables (home size, hydrocarbon potential and nitrogen oxide potential) which later OLS analyses on smaller subsets show to be of minimal importance.

Although most of the coefficient values for the variables present in both the 15-regressor and 6-regressor runs show a reasonable consistency for OLS as well as for V-ridge, on one variable, sulphur dioxide, the OLS 15-regressor equation yields a coefficient of 0.074 compared to 0.210 and 0.223 for the two 6-regressor equations.

The conclusion is that even on this data set V-ridge performs more consistently than OLS and appears to point more reliably, in the initial regression

on the whole set of predictors, to those variables which will retain their predictive value in the reduced regressions. This feature of greater consistency between initial and reduced regressions could simplify the identification of key variables in regression analysis and make the reduction process faster and more accurate.

E. Other data sets. Before summarising the conclusions from this research there are two further data sets that merit brief mention. Kerlinger and Pedhazur (1973) present as a statistical exercise a set of 20 cases in which measures of authoritarianism, dogmatism and social class are used to predict attitudes towards 'out-groups'. They conclude that social class does not make a significant contribution to the (OLS) regression of attitudes on these predictors. A comparison of the performance of OLS and V-ridge on the Kerlinger data confirms a non-significant t-figure for the OLS coefficient for social class, but shows reasonable significance for the V-ridge coefficient. In the cross-validation runs in which the coefficients from one set of 10 cases are applied to the other 10 cases, V-ridge yields an IVE of 2.30 when the samples are odd/even and 4.58 when the samples are the top and bottom halves.

Various simulation sets which are still in the process of analysis show not only that V-ridge yields results closer to the true coefficients than does OLS when there is a moderate or high level of error in the model, but also that in awkward matrices with extreme collinearity and other bizarre features, V-ridge generally stabilises the situation sufficiently to yield coefficients bearing some relation to the reality of the simulation, whereas OLS may present some coefficients which are different by orders of magnitude from the true figures, alongside other coefficients which are very close to the reality.

## 5.233 Conclusions

At this early stage in the development of V-ridge as an alternative to ordinary least squares regression it is only possible to say that V-ridge offers unusually interesting results which will need further study and the testing of a much wider range of data to determine what are the limits of performance and under what error and other conditions is V-ridge likely to prove preferable to OLS.

The following are put forward as tentative conclusions:

a. While multiple regression is a powerful technique for determining the relationship between a set of regressors and a dependent variable, the charac-

teristics of the ordinary least squares algorithm are such that there is often a gross capitalisation on chance error in the regressor matrix and an over-emphasis on the contribution made by a few variables to the criterion variance. Ridge methods offer a means of reducing the effects of error and giving recognition to the contribution of a larger set of regressors in a model. In the social sciences this wider contribution is more credible than a statistical finding that only a few variables make any meaningful contribution to the variance of some characteristic which is normally thought to be conceptually multivariate in its relationships.

b. With few exceptions, all forms of ridge regression to data have been based on the post hoc examination of the ridge trace formed by plotting the derived coefficient values against increasing values of  $k$ . This method is of research interest but cannot serve as a non-stochastic regression procedure for routine analysis of data sets. The basic assumptions for estimating the significance of the regression coefficients and for carrying out F-tests on changes in the model are invalidated by such non-stochastic procedures. The algorithm developed by Vinod (1976b) is almost the only form of ridge which offers an automatic derivation of  $k$  that is not based on the use of the criterion itself but simply on the characteristics of the regressor matrix.

c. Vinod has suggested that individual  $k_i$  additions to the diagonal of the regressor matrix may yield better results than using a constant  $k$ . Although the present programme is based on a constant  $k$ , the results from individualisation methods introduced by the present author suggest the possibility for further improvement in the V-ridge model. The methods are based on the differing variances and other parameters of the regressors and await further development. It is an area that may warrant much more research.

d. It is considered here that the use of cross-validation methods for comparing OLS and V-ridge is far more appropriate than a comparison of the performance of OLS and V-ridge coefficients on the original data from which they were derived. Comparisons based on the original data tend to favour the method which capitalises best on chance error in the regressor matrix, whereas comparisons on data sets which are increasingly different from the original data set will emphasise the method whose results have the greatest generalisability. Among the various indicators of performance yielded by the cross-validation procedures the most powerful results of V-ridge are found to occur in the area of coefficient stability and accuracy. For prediction of total variance and minimisation of residual mean square, the OLS estimators tend to yield somewhat better results, except in situations where there is a high error structure; in the latter case V-ridge performs better on almost every cross-validation indicator.

e. Statistical models which are derived from a small to medium number of cases or where a moderate to high error level occurs are likely to benefit most from an application of the V-ridge method. The potential of the method needs to be examined in fields other than that of ordinary multiple regression. In general it appears that this automated method is not only superior to the stochastic ridge trace methods for reducing the unreliability of ordinary least squares estimators, but is also likely to outperform OLS on most indicators of performance where the error levels in the data structure are at least moderate. If the evidence submitted in this study can be supported by studies on other data sets, the implications for regression analysis in the social sciences may be considerable.

#### 5.234 Application of V-ridge to present study

There were several decisions which had to be taken in the light of the extensive study and development of the V-ridge model described on the previous pages.

Firstly, rather than apply any particular rule as to the choice of OLS or V-ridge algorithms for solving each separate regression equation, it was considered more consistent to rely on a single algorithm throughout the regression analyses. In view of the considerable superiority of V-ridge, as shown in a variety of empirical and Monte Carlo trials with small to medium samples, it was decided to adhere to the V-ridge solution.

Secondly, as the V-ridge programme developed here contained three important options it was necessary to decide on a particular combination of options for the analyses in this study, again with a view to consistency in interpretation and repeatability of the analyses. To help reach a decision, two major sets of data were employed - the variables measured in the nursery and mid-test batteries. Thirteen of the nursery and 13 mid-test variables were selected for entry into separate regression equations as predictors of post-test reading attainment (a composite of the three separate reading tests used in the final battery) and of post-test mathematical concept attainment (a composite of Boehm concepts and the Piagetian test). The results from these equations were examined. In the case of the mid-test variables it was found that six variables yielded significant coefficients for both reading and mathematics attainment; for the nursery equations there were five variables which yielded significant coefficients for both forms of attainment and a further four which were significant as predictors of only one criterion. Three subsets of significant variables were made up and entered into a series of 42 regression runs in which the options were systematically varied, to determine which V-ridge options gave the highest Index of V-ridge Effectiveness (IVE) when compared with OLS results in the cross-validation previously described.

The first programme option is that of basing the regressions either on deviation or correlation matrices. The second is that of scaling the variables (for regressions based on a deviation matrix) according to histogram evidence of each variable's distribution over all 159 cases (the final size of the full sample), with each scaling being reduced or extended according to whether that variable's distribution shows the presence of outliers or whether it has a blunted distribution at one end due to skewness; clearly the alternative second option is the customary automatic scaling of variables (in the deviation matrix) to a uniform width of ten units, centred on zero. (The reasons for the equal scaling of variables in a deviation matrix have been set out in section 5.231.) Clearly the differential scaling of variables involves subjective judgements on the



shape of the histogram and the decision as to what a 'normal' range might be.

The third option is that of individualisation of the V-ridge constant  $k$ , as compared with using a uniform V-ridge  $k$  down the diagonal of the regressor matrix; the particular individualisation used here is

$$k_i = \frac{\text{Mean Var}(x_j)}{\text{Var}(x_i)} \cdot k \quad j = 1, 2, \dots, m$$

where  $k_i$  is the individualised constant for the diagonal element  $ii$   
 $k$  is the uniform V-ridge  $k$ , and  
 $\text{Mean Var}(x_j)$  is the mean variance of all  $m$  regressor variables used in the particular equation

A range of other forms of individualisation have also been developed but none of them have so far proved particularly or consistently superior.

For two of the sets of variables the runs were repeated using disattenuation procedures, to determine whether these made any difference to the comparisons.

The results (see table 17 overleaf) indicate that in only one out of the seven sets of runs does V-ridge regression based on a correlation matrix yield better IVE parameters (comparing the effectiveness of V-ridge with that of OLS) than it does when based on a deviation matrix. It is worth noting that this is not necessarily the case with all types of data; with some data sets correlation based regression is generally better, as measured by the IVE parameters. It has also been found that the more favourable IVE comparisons are often due to the relatively poorer performance of OLS with a particular matrix; this can be seen in the wider deviations between parallel sets of OLS regression coefficients when derived in the cross-validation, rather than in an especially improved performance by V-ridge.

If the main purpose of the analyses was that of maximising the superiority of V-ridge over OLS, for these data the matrix of choice would clearly have been the deviation matrix (on which to base the regressions). However the path analysis programme ultimately chosen for the analysis required the construction of standardised latent variables at various stages. This consideration and the difficult decision to base the analyses on standardised rather than unstandardised data (both issues are discussed in section 5.30) meant that it was finally decided - after carrying out extensive preliminary path analyses based on both forms of matrix - to rely on correlation matrices throughout. The repeated evidence of superior performance by V-ridge in virtually every regression on this data set, whatever form of matrix was used, and the relative equality of the total variance accounted for by V-ridge under either form, suggested that this was a reasonable decision in the circumstances.

The results of the comparison runs also indicate that for the second option,

Table 17. Trial runs to test IVE improvement with different programme options

Variable set	Criterion	Options *	Index of V-ridge effectiveness	Variance accounted for	
				V-ridge	OLS
Mid-test, 6 sign.vars. (not dis- attenuated)	Post-test composite reading	D S 0	3.59	.367	.382
		D S 1	5.01	.363	.382
		D N 0	4.01	.370	.382
		D N 1	7.06	.362	.382
		C N 0	2.75	.370	.382
		C N 1	4.35	.368	.382
Mid-test, 6 sign.vars. (disatten- uated)	Post-test composite reading	D S 0	3.91	.373	.380
		D S 1	5.25	.370	.380
		D N 0	4.49	.374	.380
		D N 1	7.85	.368	.380
		C N 0	3.33	.377	.380
		C N 1	5.24	.375	.380
Mid-test, 6 sign. vars.	Post-test composite maths concepts	D S 0	1.88	.526	.547
		D S 1	1.25	.516	.547
		D N 0	1.77	.530	.547
		D N 1	1.12	.515	.547
		C N 0	2.63	.533	.547
		C N 1	1.67	.528	.547
Nursery, 5 sign. vars	Post-test composite reading	D S 0	5.80	.480	.508
		D S 1	5.43	.474	.508
		D N 0	5.47	.476	.508
		D N 1	4.69	.463	.508
		C N 0	4.40	.485	.508
		C N 1	5.09	.475	.508
Nursery, 5 sign. vars	Post-test composite maths concepts	D S 0	3.16	.449	.454
		D S 1	2.79	.445	.454
		D N 0	3.15	.449	.454
		D N 1	2.87	.446	.454
		C N 0	2.90	.451	.454
		C N 1	2.79	.449	.454
Nursery, sep. 4 sign. vars. (not dis- attenuated)	Post-test composite maths concepts	D S 0	2.63	.407	.411
		D S 1	2.98	.407	.411
		D N 0	2.54	.407	.411
		D N 1	2.74	.409	.411
		C N 0	1.72	.405	.411
		C N 1	1.96	.407	.411
Nursery, sep. 4 sign. vars. (disatten- uated)	Post-test composite maths concepts	D S 0	2.26	.407	.409
		D S 1	2.77	.406	.409
		D N 0	2.19	.408	.409
		D N 1	2.56	.410	.409
		C N 0	1.33	.407	.409
		C N 1	1.66	.407	.409

- \* Options: (i) D regression based on deviance matrix  
C regression based on correlation matrix
- (ii) S variables subjectively scaled  
N variables automatically scaled
- (iii) 0 uniform k used in V-ridge regression  
1 k individualised on basis of programme procedure 1

Note: subjective scaling does not apply in the case of correlation matrices, for which the data are automatically centred and standardised

in four out of the seven sets of runs the subjective setting of variable ranges, as described earlier in this section, yields better IVE figures than when the variable scale ranges are based on the automatic scaling procedure. In one case the results are virtually the same and in two cases the automatically scaled data give a better comparison IVE. However in all four runs where subjective scaling is better the improvement is relatively slight; in the two runs where automatic scaling is better the improvement is quite considerable. A further consideration is that a basic goal in developing V-ridge, as opposed to most other forms of ridge regression, has been to find a non-stochastic 'automatic' ridge method which does not depend on any subjective and thus fallible judgement. The automatic nature of the method is also important with a view to repeatability. Thus, despite the small improvements which subjective scaling might bring, it was decided to adhere to the automatic scaling process throughout.

The individualisation of ridge  $k$ , which is of course a fully automatic and non-subjective procedure, does not show any considerable improvement over the use of uniform  $k$  within the present data sets; in only four out of the seven runs is the individualisation procedure superior. (Runs using the other seven forms of individualisation showed that in some cases superior results could be obtained, but these were not consistent.) As pointed out earlier in this chapter, a considerable amount of work would have to be undertaken on the mathematical concepts underlying the individualisation procedures before deciding on the conditions under which a particular procedure may yield better results than those of a uniform ridge  $k$ , within the V-ridge model.

Thus, of the three options tested here, it was finally decided to base the regressions on correlation matrices, with automatic scaling and a uniform  $k$  for any particular matrix.

Finally, several test runs established that the OLS coefficients and several other OLS parameters derived in the V-ridge programme were identical to those (OLS) results obtained when the data were analysed using the SPSS regression programme.

Ancillary material on the V-ridge programme is presented in the following appendices:

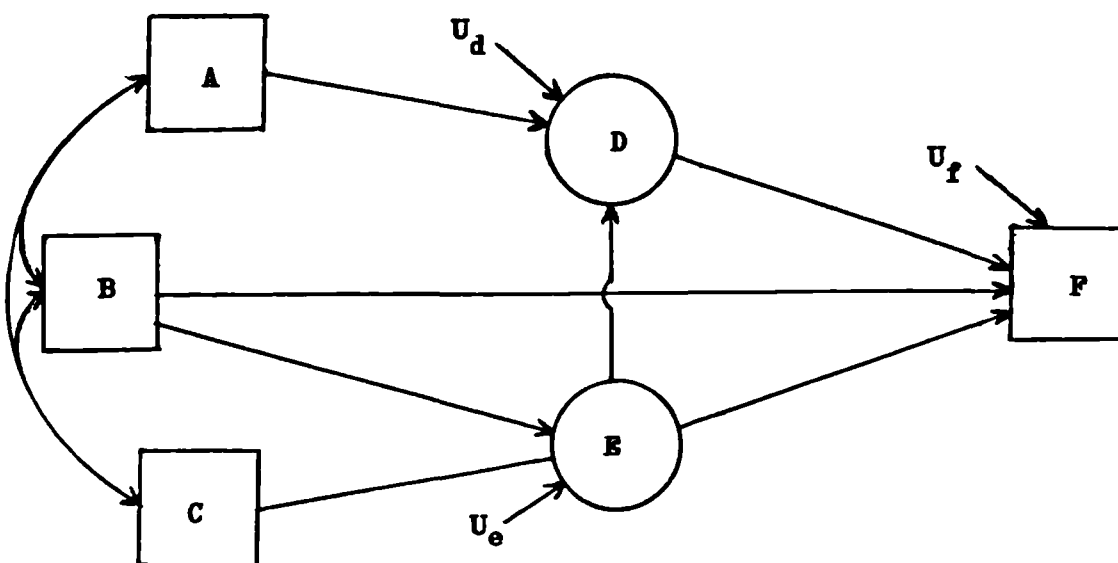
- Appendix E1: the computer programme written by the author to derive and compare Ordinary Least Squares and V-ridge regression analyses.
- Appendix E2: the V-ridge user manual for the programme in its current state of development. Further simplification is in prospect.
- Appendix E3: four typical plots of log ISRM (vertical axes) versus log  $k$ . The plots, and the accompanying programme iterations, indicate that even for grossly multicollinear matrices it is nearly always possible to identify a single lowest value of ISRM.

### 5.30 Path Analysis

Path analysis is in essence a model for analysing interrelationships between variables at a level of complexity above that of multiple regression. It is particularly suited to longitudinal models in which data is averaged at three or more points in time. It can also take account of mutual interaction between variables.

Tukey (1954), regretting his earlier failure to recognise the merits of path analysis, points out that every statistical method is to some extent the child of its own time. It is superior to correlational analysis because it can be used to examine causal hypotheses - despite Pearson's attack on it on the grounds that causation could not be proved. Tukey shows that Wright's pioneering development of path analysis, important though it was, was based on correlation coefficients; this is still a tangential method as it can only show the effect of small changes around the level of the correlation but cannot describe what would happen over a larger range of values and sample. Path analysis only becomes functional when based on regression coefficients. He points out that provided the measurements are made with relatively small errors, or alternatively that the error is taken into account by disattenuation, regression will remain reasonably constant over a wide range of situations, whereas correlations vary with every sample.

Figure 10. A simple recursive path analysis model



A relatively simple path analysis model is presented above. A, B and C are exogenous or independent variables, which may or may not relate to each other;

D and E are intermediate endogenous variables, serving both as outcome or dependent variables and as predictors or contributors to F, the final outcome variable in the model. In this case it is assumed that A and C make no direct contribution to F.  $U_d$ ,  $U_e$  and  $U_f$  are the unexplained portions of the variance of D, E and F respectively (whether these portions are due to error or unmeasured variables).

It will be seen that path analysis has the potential for assessing both the direct and indirect effects of a variable such as B. In many situations this is of considerable importance - for example, is a child's academic attainment (F) determined directly by his or her cognitive skills (B), or are those skills partly mediated by school experience (E), and if so to what extent?

### 5.301 Practical applications

Path analysis is not widely applied in research at the present time, partly because of the impression given by some authors (see later in this section) that it is an extremely complex statistical procedure. Examples of where it has been used are Peaker (1967b), who examined home and school variables in predicting to school test scores for the Plowden Report (H.M.S.O., 1967); Coleman (1975), whose sensitive use of path analysis showed the importance of home variables in educational outcome at 10 years, in contrast to the IEA (International Education Association) conclusions from the same data that school variables are much more important than home variables; Hauser (1971), who developed sophisticated models relating socio-economic, cognitive and motivational variables to each other and to school performance; Goldberg (1971), who used path analysis to test a variety of alternative explanations for voting behaviour; Duncan and Featherman (1973), who assembled a number of social, psychological and cultural variables to derive a path model with education as an intermediate and occupation as a major outcome; and Halsey, Heath and Ridge (1980), whose conclusions about family and child characteristics and social class as predictors of various educational outcomes and decisions (e.g. on school-leaving age) rested largely on path analyses.

As Goldstein (1976c) shows, interest is developing in the application of path analyses to longitudinal data, rather than relating measurements at different points to time and age variables. But he warns of the serious problems of measurement unreliability which have to be overcome or allowed for in the analyses, and the identification problem in that it may be impossible to estimate all the parameters of the equations uniquely. He refers to the application of path analysis to log-linear models as a more recent development permitting the treatment of proportions in the same way as continuous measurements.

5.302 Theoretical development

As Tukey (ibid) has recognised it was Wright (1921) who first developed the basic concept of relating variables statistically within a path system, "and thus finding the degree to which variations of a given effect is determined by each particular cause". Wright relied mainly on correlation coefficients and what he termed 'path regression coefficients' (essentially partial regression coefficients). The equations from which the parameters of a model were derived were essentially of the form

$$r_{fe} = p_{fe} + p_{fb} \cdot r_{be} + p_{fd} \cdot p_{de}$$

where  $r_{ij}$  are the correlations between variables  $i, j$

$p_{ij}$  are the 'path regression coefficients', and

$b, d, e, f$  apply to the variables in the path model portrayed  
two pages previously

Given the correlation coefficients relating a set of variables to each other, it is usually possible to construct sufficient equations to solve a system of relationships (unless the system is underidentified, so that the number of paths is greater than the number of possible equations).

Despite the cautions voiced by Tukey (1954) about this approach, Wright (1960b), Duncan (1966) and many of the authors who contribute to Blalock (1971) and to Goldberger and Duncan (1973) build their models upon the solution of a series of equations which seek to identify the component parts of correlations between the variables in the system. There is no doubt that this is a sensitive approach when one is examining each element of a model in great detail and seeking to establish a pattern of interrelationships such as in Coleman (1975) or Halsey (1980). On the other hand if one is deriving a predictive model in which the reliability of the derived coefficients is of particular importance it appears that direct regression equations offer a more satisfactory solution because of the considerable variability of correlation coefficients. Even the relative instability of regression coefficients cannot compare with the problems arising from correlational instability. It may also be argued that the correlational approach, as reflected by the variety of contributors to Blalock (1971) and elsewhere, offers an unnecessarily complex approach to any form of path interpretation.

A work of considerable value which examines a variety of path models - recursive and non-recursive - is that of Namboodiri et al (1975). The authors see the main source of deviations from prediction as being due to specification errors, including random and non-random measurement errors, omission of relevant variables from the model, and errors in equations due to non-linearity or non-additivity. Errors such as these make significance tests meaningless, in their

view. Another problem is that omitting paths which do not reach a certain level of significance can leave a model that is more likely to be rejected in an overall test of significance; on the other hand the continued addition of paths could yield a solution which would satisfy more than one model. The authors interpret a path coefficient as a standardised regression coefficient.

Other useful theoretical studies in this field are those of Land (1969), who examines the correlational approach to path analysis in great detail; and Macdonald (1977), who points to the many problems and doubts that arise in the application of path analysis. Macdonald rejects the 'conventional' approach of correlational decomposition and suggests rather the use of least squares - which emphasises the similarity between this and other well established statistical techniques. He considers that even binary variables can be invaluable and blameless as predictors in a path model. Ordinal variables can also be treated in this model as though they are interval variables. He discusses methods of deriving indirect effects and of eliminating weak or negligible predictors (relying on the *t* statistic for the derived coefficients).

Earlier sections of this study have examined briefly the reasons why alternative methods of analysing the data were not given preference, such as multivariate analysis of variance (with or without controlling for covariates), canonical correlation and factor analysis. Despite their various merits none of them can offer a direct interpretive model of multiple interrelationships across a set of variables, with useful though not infallible criteria for assessing the strength of those relationships.

Path analysis not only provides a means of achieving this goal but also offers a particularly insightful method of examining relationships longitudinally. The importance of the longitudinal examination of relationships in the field of development has been studied by Hindley (1960), who points to the advantages and limitations of such studies. Wall and Williams (1970) review a wide range of studies and conclude that future studies will need to sample in such a way as to determine the effects of environment on specified groups of individuals. Both Hindley and Wall and Williams emphasise the shortcomings of cross-sectional analyses for interpreting development. Gallagher et al (1976), in their examination of longitudinal research design in relation to development, show how it is virtually impossible to relate maternal environment to later child outcome, for example, unless the data is analysed longitudinally. The problems of longitudinal studies, dealt with by all the authors cited in this paragraph, include the retention of samples, alternations in programme or secular history changes between the initial and final stages of the research, practice effects in the instruments used, and the work and expense involved in following samples over time.

It is difficulties such as these which have limited the use of longitudinal

research in the past. Yet the answers to important developmental hypotheses can only be obtained painstakingly over time, and analysed similarly; the post hoc collection of data to examine such hypotheses is always a second best solution.

If longitudinal research has a particular justification for examining change over age and time, the statistical methods for analysing the data should be appropriate to the nature of the research. Path analysis is probably the only viable method and possibly the most sensitive, flexible and specific way of presenting and testing longitudinal hypotheses. The extra effort and the longer time span demanded for the analysis of this type of research may well be justified by the greater insights achieved.

### 5.303 Specific approaches

The development of statistical algorithms for dealing with path models can be examined in relation to three disciplines - sociometrics, econometrics and psychometrics.

Sociological studies have produced by far the largest number of algorithms for handling the analysis of interrelationships over time; the sociological literature has many examples of the practical application of different path models.

Boudon (1968), arguing within the correlational approach to path analysis, offers 'dependence coefficients' which are corrected by the appropriate variances of the related variables. Costner (1969) stresses the need for two or even three indicators of each theoretical variable in a model; however he uses correlational rather than regression techniques to derive the theoretical variables. Blalock (in 1972 and other studies previously cited) deals with the intricate problem of causal interpretation and focuses on the varying interpretations that are possible even with only three interrelated variables. Goodman (1973) offers tests for a whole system of equations and looks in particular at panel studies. Another fine-grained theoretical study is offered by Sullivan (1974); he looks especially at the problem of error in variables and also concludes that multiple indicators are useful in deriving theoretical variables.

It is Duncan (1966, 1969, 1971 - adding some critical comments to the 1966 algorithms, and with a very clear exposition of the correlational approach on pages 137-8 - and 1975), Goldberger and Duncan (1973) and Goldberger (1973) who offer a most comprehensive examination of specific theoretical approaches to path analysis within the sociometric field. Duncan's reviews are highly competent, fully aware of the shortcomings endemic in most approaches to path analysis, and yet recognising the value of such analysis. After examining the problems of interpreting a variable such as motivation within a system of relationships over time



he concludes that an analysis based on theoretical variables will differ considerably from one based on naive analysis. In his 1975 study Duncan recognises that many path analysis models in the sociological literature are flawed because of measurement error and too facile assumptions on disturbances; any model has to rest on sound theory within its own discipline. He considers that indicators of covariance do not provide evidence for a model; one needs proper theory to invent hypothesised variables and then to test and confirm these with a structural model. He discusses the theory of structural equations in some detail, as do Goldberger and Duncan (ibid); the latter point to two-stage least squares as an improvement on simple least squares regression for path analysis.

The field of econometrics has also made important contributions to the development of path analysis algorithms. Early workers such as Haavelmo (1943), and more recent studies by leading econometricians such as Theil (1971), Johnston (1972), Klein (1974) and Koutsoyiannis (1977) present in depth a variety of methods for handling structural equations. In essence these are multiple regression equations which in combination define the various interrelationships within a system. Referring again to the model in figure 10, one might derive the set of recursive equations (recursive implying that any variable in the path sequence from left to right can only affect a later variable):

$$\begin{aligned}\underline{f} &= p_{fb} \cdot \underline{b} + p_{fe} \cdot \underline{e} + p_{fd} \cdot \underline{d} + u_f \\ \underline{e} &= p_{eb} \cdot \underline{b} + p_{ec} \cdot \underline{c} + u_e \\ \underline{d} &= p_{da} \cdot \underline{a} + p_{de} \cdot \underline{e} + u_d\end{aligned}$$

where  $\underline{a}, \underline{b}, \dots, \underline{f}$  are the individual (standardised) values of the variables A, B, .... F

$p_{fb}, p_{fe}, \dots, p_{de}$  are the ordinary regression coefficients relating the outcome variable (first subscript) to the predictors, and

$u_d, u_e$  and  $u_f$  are the disturbance terms or unexplained contributions to each outcome variable

The addition of constants would clearly enable the use of non-standardised variables in this model.

With two unknowns (variables D and E) this set of simultaneous equations is easily resolved. However it is not difficult to envisage a more complex interrelationship between a set of variables in which a system is overidentified - more than one solution to a set of equations being possible - or alternatively underidentified so that no solution can be derived. For the underidentified model it is essential to obtain fuller data to reduce the number of unknowns or alternatively to make simplifying and theoretically credible assumptions to enable a tentative solution to be derived.

It is the overidentified system whose uncertainties have led to major developments in the statistical solution of path models. Overidentified systems occur particularly in non-recursive systems (systems in which some of the relationships between endogenous variables are reciprocal or where particular relationships are in a reverse direction to that of the remaining relationships surrounding that variable in the model).

Theil (ibid) and some of the other econometricians cited earlier have developed the theory of two-stage and three-stage least squares to a high degree to deal with sets of over-identified structural equations. Essentially this involves starting with 'reduced form' equations in which each endogenous variable is regressed on all its predetermined or independent variables to derive an initial estimate of that variable. These estimates are then used in the complete system of equations to obtain a solution.

Three-stage differs from two-stage least squares in several respects - principally in that the former adds a third stage of generalised least squares which takes account of restrictions such as covariance between the disturbances of the different equations. To that extent the three-stage algorithm is closer to the reality of a system of variables where the assumption of uncorrelated disturbances is seldom realised. On the other hand the three-stage requirement that the complete specification of the entire system should be known, including the mathematical form of that relationship (Koutsoyiannis, 1977), is a requirement which is generally too demanding, certainly within the social sciences.

In contrast to the field of sociometrics, where Duncan (1969 and 1975) warns against the incautious use of path analysis, psychometrics appears to have been over cautious in its approach to this form of analysis. Path models are seldom developed to handle the complex interrelationships of the large number of variables assessed in major research endeavours. Combined educational and psychological studies have occasionally attempted the solution of structural models - Peaker (1967b), Coleman et al (1966) and Jencks (1972), for example - but more generally the analyses of large sets of variables have been limited to analyses of separate subsets of the data, relying on covariance techniques to control for single variables such as social class or 'intelligence'.

The contributions of Werts and Linn (1970), Werts Jöreskog and Linn (1973), Werts, Linn and Jöreskog (1974), Jöreskog (1976) and Jöreskog and Sörbom (1976) are focused on models having a very small number of variables in which complex procedures are followed to estimate error structures and reduce the influence of such error. The reality of an educational or psychological model in which a considerable number of influences are competing in their contributions to school or developmental outcomes is totally ignored by such fine-grained analyses, which belong more to the psychological laboratory or the atypical 'controlled classroom'

situation. What should have been the starting point of path analysis for psychometrics has up to now remained its summation.

There are good reasons which might be advanced for this hesitation - reasons concerned with the disciplines themselves rather than with the statistical expertise within a particular discipline. Conclusions reached on the basis of a sociological path model may add to theory and in the long term may influence the thinking of sociologists and ultimately help to mould social policy. The study by Halsey, Heath and Ridge (1980) helps point to social and family factors which may influence later educational decisions for the child; the conclusions are not threatening nor do they have immediate policy consequences. In econometrics the conclusions of path models are often of immediate relevance to policy but this is acceptable in a field where risk-taking is integral to the decisions made on the basis of analyses of past results and prediction of future trends.

In educational and psychological research the results of path analyses can have far-reaching consequences for school policy in the short and long term and for decisions concerning the individual child or adult; thus there has been, understandably, a greater caution about positing complex models whose statistical limitations are well known to researchers. An example of the critical response by ten educational psychologists to the path model advanced by Jencks (1972) appeared in the Harvard Educational Review (Edmonds et al, 1973). Jencks had posited a model which he claimed showed that schools made almost no difference to life outcomes such as income and that it was largely social factors and 'luck' which predicted to high income.

Edmonds et al focused their attack not only on the educational evidence contradicting the Jencks hypothesis but also on the analysis itself. "... The statistical technique used, path analysis, ignores any possible non-linear relationships that might exist between family background, schooling, and success; it ignores any 'interactive' or 'conditional' relationships that might have been present in the data; it ignores any 'two-way' causal relationships (for example, Jencks examines only the effects of education on occupation, not the possible effects of ones occupation upon later education); and it ignores all variables or factors in an individual's past life or school environment that would not constitute what statisticians call 'interval scales'. These shortcomings could well result in an underestimation of the true effects of schooling and related factors on later life success."

Whether or not these criticisms are justified - it does appear that Jencks' model was far from adequately specified in terms of input variables and reciprocal relationships - the fact is that model misspecification or the presence of gross error in measurement, with other problems previously referred to, can have much more serious long term consequences for educational policy than can similar fail-

ings in a sociological model. The grave shortcomings in the initial analyses of Headstart (chapters 2 and 3) had long-term consequences in reducing the funds allocated for this form of aid. It is worth noting, however, that the analyses in question were not path models but covariance analyses in which some of the basic assumptions of covariance were ignored.

The fact that path analysis is open to errors of model specification and of variable measurement is not however a justification for its being seen by some as a rather esoteric procedure or a procedure too fallible to be trusted. There is no statistical model which is not open to model misspecification and measurement error. Very simple models based on only a few variables may give an extremely limited or totally misleading explanation of relationships. Error is always present and needs to be taken into account in any model. Path analysis can at least claim to attempt to include a considerable variety of contributors to educational or psychological outcomes and to that extent may offer a more credible model than any simpler presentation.

It remains only to cite two highly competent statistical treatments of path analysis. Macdonald (1977) offers a critical but valuable statistical description of the potential and pitfalls of path analysis. He ends with a caution that while a path model can show how individuals are responding within a particular system, there are dangers in extrapolating from this model to the consequences of a change in policy and thus in the model or system itself. However, Macdonald's brief is to describe path analysis as applied to post-hoc social survey research; he does not deal with the use of path models to explore an intervention variable which is itself a contributor to the system, as in the research which is the subject of the present study.

A second and more extended treatment is presented in Namboodiri, Carter and Blalock (1975), who go into considerable detail on the handling of recursive and non-recursive structural models, including the use of lagged endogenous variables. They describe a method of theory building based on block-recursive models - models in which 'blocks' of variables are created to simplify, conceptually and statistically, an analysis in which there are a large number of exogenous or lagged variables. Groups of exogenous variables which are highly interrelated (or even having reciprocal causation) within a particular conceptual framework can be formed into a combined index which then becomes part of the wider model of relationships between blocks. Likewise exogenous variables which are assumed to affect directly only a relatively limited group of endogenous variables can be combined within a block and then related to the endogenous variables concerned. The importance of this approach is referred to again in sub-section 5.305.

### 5.304 Standardisation

The issue of whether to base path analysis on standardised or unstandardised variables is a difficult one to resolve since there are arguments in favour of each method.

In general most of the major authors find more arguments in favour of unstandardised variables. Blalock (1971) states that the interpretation of causality demands unstandardised coefficients, with standardisation restricted to situations where descriptive analysis is required. Werts, Linn and Jöreskog (1974) point out that standardisation is unsatisfactory for growth studies since it ignores changes in true variance over time; in a typical achievement study the true variance will increase over time. Macdonald (1977) points to the difference in interpretation between saying what is the expected effect on the predictor variable of a one standard deviation change in an independent variable, and examining the effect on the outcome variable (a reading score, for example) of a specific change in the independent variable (such as the length of schooling or hours of instruction); the latter clearly represents the unstandardised model. He emphasises that it is possible to mingle standardised and unstandardised variables in one path model, provided the same scaling is adhered to throughout the analysis. Namboodiri et al (1975) show that although the difference between the two forms of analysis is seemingly trivial, the comparison being that of

$$\begin{aligned}
 y &= b_0 + b_1x_1 + b_2x_2 + \dots + e, \\
 \text{with } \frac{y - \bar{y}}{s_y} &= \frac{b_1 \cdot s_{x_1}}{s_y} \cdot \frac{x_1 - \bar{x}_1}{s_{x_1}} + \frac{b_2 \cdot s_{x_2}}{s_y} \cdot \frac{x_2 - \bar{x}_2}{s_{x_2}} + \dots + \epsilon \\
 &= B_1 \frac{(x_1 - \bar{x}_1)}{s_{x_1}} + B_2 \frac{(x_2 - \bar{x}_2)}{s_{x_2}} + \dots + \epsilon,
 \end{aligned}$$

where  $y$  is the dependent variable  
 $x_1, x_2 \dots$  are the independent variables  
 $\bar{y}$  and  $\bar{x}_i$  are the respective means  
 $s_y$  and  $s_{x_i}$  are the respective standard deviations  
 $b_0$  is the constant in the unstandardised equation  
 $b_1, b_2 \dots$  are the unstandardised regression coefficients  
 $B_1, B_2 \dots$  are the equivalent standardised coefficients, and  
 $e$  and  $\epsilon$  are the respective disturbances for the two equations,

the interpretation of results from each form differs considerably. Standardised coefficients are population specific, so that one can generalise from a sample to

its population; unstandardised coefficients enable generalisations to be hypothesised across populations or within the same population over time.

Hauser (1971), in his study of the influence of sociological and school variables on educational performance, faced this problem of which form of path analysis to use. While he recognised that changes in 'policy variables' are more easily grasped in raw form, the nature of his research suggested using standardised measures. The fact that they may change from population to population does not mean that they must change; on the contrary, it is possible to argue that interaction effects may vary across populations so that the metric measures of effect will not be invariant. Furthermore, the strength of the relationship between variables in a social system (implicitly requiring standardisation for comparative assessment) may be as interesting and important as the exact function describing such relationships. Finally, according to Hauser, the utility of metric (raw) coefficients is limited if direct measurements are not replicated or if unobservable constructs (endogenous variables) are derived and given theoretical interpretation.

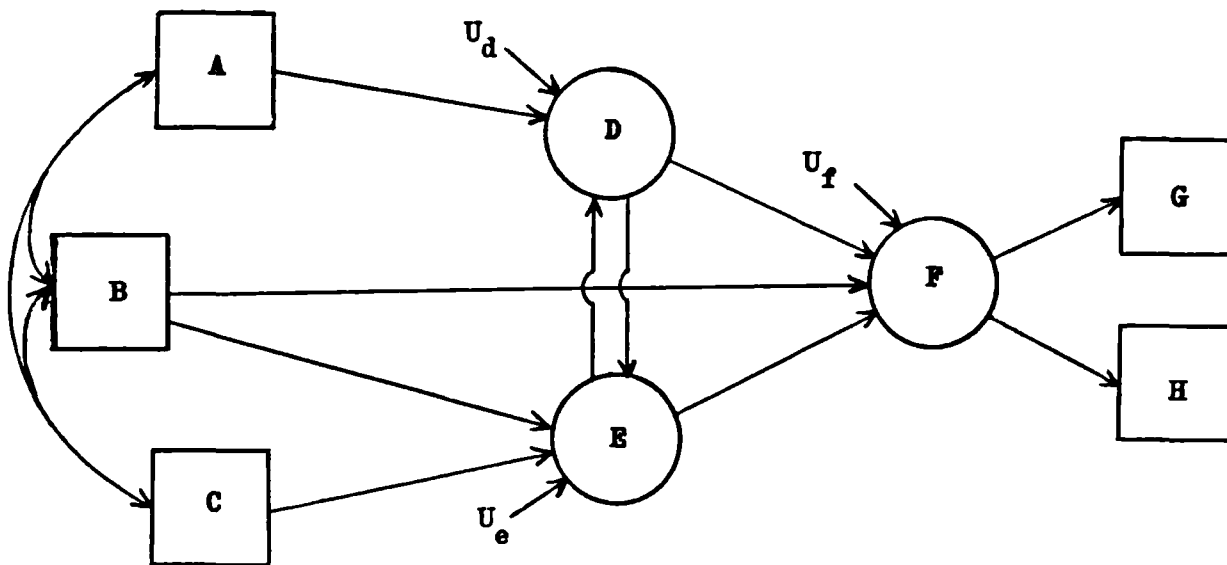
#### 5.305 Choice of algorithm

None of the major algorithms within the field of path analysis seemed entirely appropriate to the present study. The most desirable approach of applying two-stage least squares to an interpretive model in which reciprocal interactions between variables were taken into account would have required a highly complex set of equations to cope with the many variables (over 50) likely to be entered into each set of simultaneous equations; it would also have demanded the development of a further computer programme to solve each set of equations in terms of the V-ridge algorithm rather than the usual ordinary least squares algorithm.

While this possibility was being considered a model developed by Wold (1975) and tested by Noonan and Wold (1977) came to notice. Herman Wold, a leading statistical theoretician, addressed the Fourth Berkeley Symposium on Mathematical Statistics and Probability (Wold, 1961) on the derivation and characteristics of unbiased predictors and the extent to which the direction of causal relations can be assessed by purely statistical devices. In the 1961 paper he emphasised that unbiased predictors, not being reversible, placed on the model builder the burden of designing part or the whole of the model in terms of directed (asymmetric) relations.

In two other papers, Mosbaek and Wold (1970) and Wold (1975), this asymmetric model was developed further within what came to be known as the iterative fix-point method for estimating the parameters of simultaneous equation models.

Figure 11. A non-recursive path analysis model



Among the advantages seen for the fix-point method are that it is more flexible in handling non-linear relationships between endogenous variables and that its iterative feature "squeezes the data for more information than two-stage least squares", with which the fix-point method is principally compared. The full title of the method is Non-linear Iterative Partial Least Squares (NIPALS). Wold (ibid) sees the method as an evolution of causal and predictive analysis within path models, merging the multi-relational forecasting of econometrics with the modelling through latent variables characteristic of psychology and education.

The importance of this approach can be understood by reference to the simple model illustrated earlier and presented above in slightly modified form. G and H are measured outcome variables which are hypothesised to represent the latent attainment variable F. Reciprocal interaction is also portrayed between latent variables D and E.

A two-stage least squares solution to this model attempts to solve a set of simultaneous equations defining all the relationships within the model. Thus directionality and prediction are not the prime focus of the two-stage solution but rather the maximisation of all the relationships. In contrast the fix-point solution attempts to centre the solution on the hypothesised latent outcome F, doing this via an iterative method which will be described briefly below. In essence the approaches differ in that the two-stage solution seeks the best explanation of the relationships while the fix-point solution seeks a model in which explanation and prediction are combined.

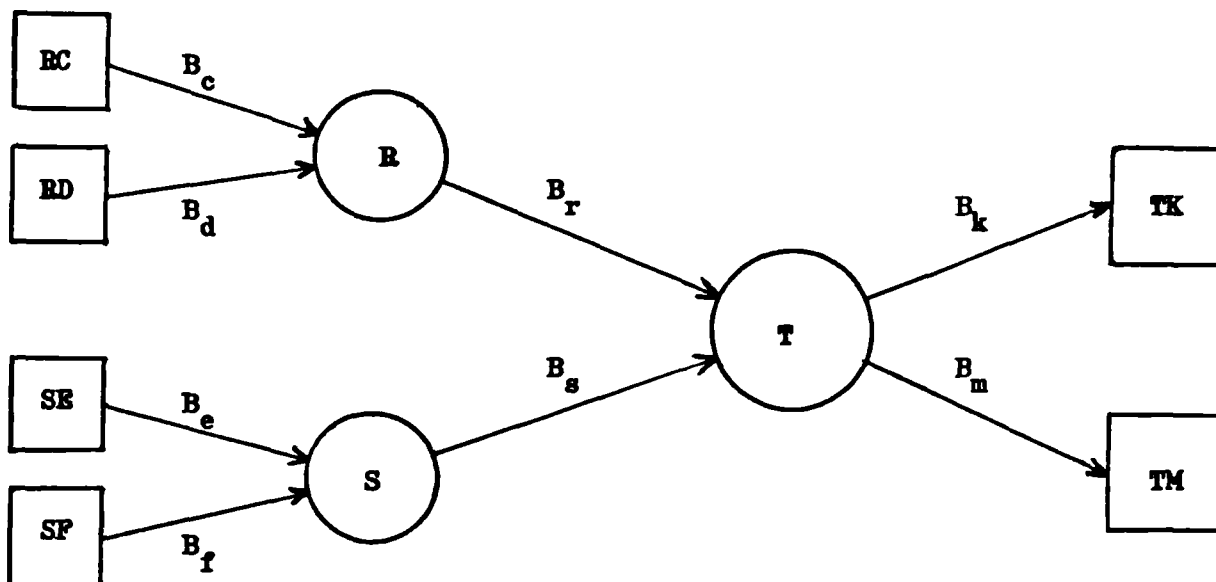
Namboordiri et al (1975) point out that the NIPALS model requires fewer assumptions about the covariances of the disturbance terms and the exogenous

variables than do explanatory models such as those of Hauser and Goldberger (1971), although NIPALS is not as efficient as the latter if the model is correctly specified. In other words, NIPALS may be more useful where there is uncertainty as to the correct specifications and other assumptions of an exploratory model. Klein (1974) also writes in praise of NIPALS as a simple and yet full information method, taking account of all the restrictions imposed on a complete system although it makes no use of error covariance information.

The theoretical justification for the NIPALS approach for handling path models with latent variables is set out in detail in Wold (1975). Predictors are estimated by ordinary least squares where all the variables are directly observed, and by iterative OLS where latent variables are involved. If all variables in the model are directly observed the solution is a straightforward application of regression of the outcome variable on to the predictor variables. If latent variables are conceptualised then iterative regressions are undertaken to maximise the correspondence between the observed outcome variables (G and H in the diagram on the previous page) and the exogenous predictors A, B and C.

Noonan and Wold (1977) set out the details of the iterative technique.

Figure 12. Application of the NIPALS method to a path model



In this method the latent outcome  $T$  (see above diagram) is initially assumed to be a composite  $T_1$  of the measured outcome variables  $TK$  and  $TM$ , with coefficients  $B_k$  and  $B_m$  initially equal to unity. The path (regression) coefficients  $B_r$  and  $B_s$  are likewise assumed to start at unity. The newly created  $T_1$  is now regressed on to variables  $RC$  and  $RD$ , and  $T_1$  is likewise regressed on to variables  $SE$  and  $SF$ . The coefficients derived in each case ( $B_c$  and  $B_d$ ,  $B_e$  and  $B_f$ ) are used to create the latent variables  $R_1$  and  $S_1$  respectively.  $TK$  and  $TM$  are then regressed on to the composite  $(B_r R_1 + B_s S_1)$  to derive new values for the coefficients



$B_k$  and  $B_m$ . These new values are used to create  $T_2$ .  $T_2$  is now regressed on to  $R_1$  and  $S_1$  to derive new values for  $B_r$  and  $B_s$ . The whole cycle is then repeated, regressing  $T_2$  on to BC and RD and likewise regressing  $T_1$  on to SE and SF to derive new values for  $B_c$ ,  $B_d$ ,  $B_e$  and  $B_f$ . The procedure continues as already described, iterating until a satisfactory degree of convergence occurs between successive values of the main path coefficients. At each individual stage within the cycles the latent variables are standardised. While the Noonan and Wold (1977) example also specifies the use of standardised exogenous variables, this is done without loss of generality.

There were a number of reasons why it was decided to adopt the broad principle of the NIPALS method for analysis of the path models in the present study:

- a. The approach of Wold, whose method combines overall predictive power with the modelling of causal hypotheses, is similar to the general approach of the present study, in which prediction of the outcome scores in reading and mathematics is seen to be as important as the interpretation of the relationships between the large number of variables assessed in this study.
- b. The NIPALS method specifically focuses on the relationship between the outcome (attainment) variables and each of the large number of exogenous contributor variables, while its openness to the grouping of these contributors into 'block recursive' latent variables is an equally desirable feature. Without this grouping the models as a whole would be limited to the most powerful individual predictors or to those variables whose position along minor axes of the variable space enabled them to remain in the models.
- c. The procedure of grouping individual variables (by using the NIPALS regression procedures) into stronger latent variables may help to minimise the tipping effect referred to by Gordon (1968) in his important study of the instability of regression coefficients.
- d. An extension in this study of the thinking underlying the NIPALS model enables the inclusion of successive latent variables in an extended longitudinal model; in this extension the latent variables derived from the pre-test variable sets also serve as ordinary contributory predictors of the mid-test latent variables, and these in turn serve as predictors of the final (post-test) attainment outcomes.
- e. The NIPALS method is more open to the application of the V-ridge regression technique developed in the present study than is two-stage least squares or other solutions to structural equation sets.

While the nature of the path models developed in this study is more complex than that described in the Noonan and Wold (ibid) explanatory example, the estimation procedure proved to be considerably easier than expected since iteration

was not required. Extensive proving runs, using iteration, suggested that the outcome variables - whether seen as a conceptual 'reading' outcome based on the scores of three different reading tests, as a composite 'mathematics' outcome or as a hypothesised 'total attainment' combining various test scores - should preferably not be subjected to iterative change, as would have been the case had it been the intention simply to maximise prediction between the contributor variables and any possible combination of outcome scores.

What occurred in the proving runs was that the iterative process tended to maximise the importance of certain outcome variables and minimise other outcomes within the hypothesised latent variables. For conceptual reasons however it was felt that a specific composite of variables, in proportions decided in advance, should be defined as representing each latent attainment; these specified combinations were total attainment, reading attainment, and two types of mathematics attainment. While iteration would maximise the predictive power of the model it would also sacrifice the precise definition of each latent outcome.

For example, it was considered essential to posit 'total attainment' as a particular combination of the three reading, two mathematics and Piagetian scores. Iterative runs on the data, as in the full NIPALS method, with a flexible outcome variable, showed that the attainment combination was weighted increasingly heavily in each round by reading - as could be expected with a child population whose age at post-testing was between 5 and 6 years. Since the research hypothesised goals of increased attainment in all three areas of reading, mathematics and Piagetian skills, it was essential to state in advance how these outcomes were to be combined within a single latent variable for the path models based on that outcome. The same principle applied to the formation of the other latent variables.

In the present study it was also decided to rely on standardised data, particularly as there is a strong focus on assessing the strength of competing variables at each time point or stage in the models. While it may have been interesting to assess the influence of the parent programmes in terms of sessions attended, the fact that the programme variables contribute both directly and indirectly to the ultimate outcome variables suggests that it may have proved impossible to interpret the raw programme attendance 'influence' in any meaningful way. The alternative method followed, in which 'parent programmes' were standardised like all other variables, enabled the specific contributions to be weighed in clear competition with the contributions of other variables.

#### 5.306 Disattenuation of all matrices

In all the path models derived in this study each separate regression matrix

was disattenuated to take account of the differing reliabilities of each predictor variable. As described earlier, in section 5.14, disattenuation is applied to the matrix of correlations between the set of regressor variables, prior to inversion of the matrix and the subsequent derivation of the regression coefficients.

When a latent variable is derived from a combination of prior variables its own unreliability is determined on the basis of the algorithm

$$\frac{\sum (r_{ii} \cdot B_i)}{\sum B_i} \quad i = 1, 2, \dots m$$

where  $r_{ii}$  is the reliability coefficient of each contributor variable, and  
 $B_i$  is the standardised regression coefficient for each variable

This unreliability figure is used for disattenuation in any matrix in which the latent variable appears as a predictor.

#### 5.307 Data-handling programme

In view of the fairly large number of data variables obtained in the course of this study (some 60 raw variables, 8 of these being categorical, and a considerable number of derived variables such as standardised measures and latent variables made up from subsets of raw variables) it was necessary to handle the data on computer in an efficient manner if the planned volume of path analyses was to be undertaken.

The data itself was placed on computer files and a special file manipulation programme was written (in Fortran, as was the original V-ridge programme) to carry out a range of basic operations with a minimum of manual input work. This programme, termed Remul, is printed in full in Appendices E4-5.

The programme is initiated with a relatively small number of cards which specify the file sources and formats of the particular variables required for a particular analysis, such as multiple regression. Up to eight different files can be accessed, seven of them being used within any single run of this programme; each file can contain up to 17 variables. The output can be placed in any one of three files prior to further analysis of the prepared variables.

There are five forms of variable manipulation which can be undertaken. Variables can be assembled from any or all of seven computer files at one time and output on to a single file; the output can be standardised, if required, yielding the mean and variance of each variable; groups of cases can be identified and selected out for separate analysis (such as boys, girls, school or

ethnic groups); a subset of cases can be selected on the basis that the scores on one of the variables are above a specified value and/or below a second specified value; finally the scores of a latent or 'predicted' variable can be derived using a constant and regression coefficients appropriate to a particular set of variables.

Operation of the programme is designed to be simple and straightforward, with several built-in checks against incorrect entry or specification of the data, in so far as these errors can be identified by the programme itself.

## 6.00      R e s u l t s

The results of the study are set out in detail in this chapter.

The nature and quality of the tests carried out on the children, to determine whether the reading and mathematics programmes given to their parents are effective or not, are of fundamental importance in the assessment of the intervention. Section 6.10 reports on these tests and on the numbers tested at different stages, as well as on the problems encountered. The efforts to retain as much of the sample as possible are set out; an examination of the attrition sample shows that the differences between that group and the surviving sample are not critical.

While the child tests serve to assess the cognitive characteristics and early attainment of the sample children, the level of stimulation or otherwise in the parent environment is of equal importance in assessing the influences on a child's performance. Thus section 6.20 deals with the parent interviews, including the parents' description of the reading and mathematical 'environment' in the homes, their attitudes towards school and school work, and their comments on nursery education. Other issues discussed in this section include television viewing in the homes. The important procedure for allocating parents to experimental groups is also reported here.

The intervention itself took the form of parent reading or mathematical programmes, given to parent groups in rooms adjoining the nursery classes. The meetings took place once a fortnight over a period of four months. A large number of small groups were set up in the six sample schools. Section 6.30 reviews the goals and actual work in these programmes, including the pattern of attendance; the parents' views on the programmes, obtained at the end of the intervention period, are described at some length. Methodological aspects of this work are also examined.

Section 6.40 examines the nature and characteristics of the data themselves. Despite the recognised limitations of what are termed 'soft' data in the social sciences, there is enough evidence that these data offer a reasonable foundation for use in the planned analyses. The various divisions of the sample are also examined, including the divisions based on sex, social and ethnic categories, and groupings based on allocation to the different parent programmes. The characteristics of all these divisions are examined briefly. An important sub-section presents a commentary on the experimental design and planned analysis.

Two issues of crucial importance are reviewed in the next section (6.50), namely the reliability and validity of the data. While the determination of reliabilities for the test data was relatively straightforward, the problems of obtaining reliability figures for some of the other measures were more complex. Validity is approached mainly from the viewpoint of the newly developed indices of nomological validity and nomological redundancy. The characteristics of all the data are examined, in regard to both reliability and validity.

A particular methodological problem in the study was that of assessing the influence of the different age variables, since it was not possible to carry out all the tests at a particular age level, nor could the children themselves be selected from within a narrow age band. Rather than try to correct scores for age of administration - with the imponderable issue of how varying age patterns within schools and across the sample might bias those corrections

- the various age variables were used as predictors in their own right. This yielded a number of useful insights. An even more complex methodological problem arose with the assessment of the influence of the length of time that any particular child spent in the Nursery class, and subsequently in the Reception class. After examining a variety of alternatives a particular formulation was devised which appears to offer important evidence on the academic value or otherwise of the period spent in the Nursery and Reception classes, for different groups. The age and time issues are discussed in section 6.60.

The use of correlations and group scores in examining the 'success' or 'failure' of the study is discussed in section 6.70. It is shown that reliance on such indicators in a sample having many different characteristics is almost impossible, given the range of influences operating on the children in addition to the influence of the parent programmes.

The ultimate goal of the analyses has been the development of path models to interpret the relationships within the data and to provide evidence on the effectiveness or otherwise of the parent programmes. Section 6.80 sets out the rationale for using a variety of statistical techniques within the path models and lists the various criteria used, both in constructing the models and in accepting or rejecting particular paths. Each of the ten models is then dealt with in turn, accompanied by diagrams of the initial creation of latent variables and the final path models themselves. The tentative findings from the models are discussed and particular conceptual and statistical problems are examined in some detail.

## 6.10 Child tests

The tests were carried out according to the protocol outlined in sections 4.31 and 4.34, and in the Appendices A.

Testing started at the pilot school in late October 1976 and various minor adjustments were made in this period, including the decision to reject one test (of social awareness) and drastically alter another (distractibility) because of serious reliability problems. Tests then started at the first of the six sample schools in mid-November and continued until the middle of February. Children of 4 and over were administered the full nursery battery. The minority of children who were not yet 4 at the date of testing or who were judged to be rather immature were given the four basic tests of 'attainment', namely the E.P.V.T., reading awareness, the Infant Reading Test and the mathematics test. This was considered necessary in view of the impending start of the parent programmes, which it was hoped would influence the children's levels of pre-reading and pre-mathematics. The remaining cognitive tests were given to the young children later, usually within a month or two of their attaining the age of 4.

As the sample children moved up into reception class and started full-time education, the battery of reception class tests was administered. This was usually done about a month after entry. The reception battery was likewise tried out in advance at the pilot school. For organisational reasons (including the concurrent running of the parent programmes) it was not always possible to abide by this schedule of testing one month after entry, and in a minority of cases (about one-seventh of the sample) the reception tests were only administered several months after the school entry of children who started in the first half of 1977. The possible effect of this on later analyses would be that the statistical influence of the reception class could be slightly reduced and that of the nursery class equivalently increased, in one-seventh of the sample, but even this small effect would only be an indirect one across the total framework. As the reception battery involves only cognitive tests, the effect on any conclusions about the factors contributing to reading and mathematics progress would be so minimal that it can safely be disregarded.

Because of the 19-month age span of the sample children, as described in section 4.2, reception class entry ran from January 1977 to Easter 1978. The final reception tests were however carried out in late January and early February 1978, on all the remaining sample children (one-quarter of the sample). Of this total of about 35 children, some 10 were only due to move up from nursery to reception class at Easter 1978. Thus these last few children were tested while still in their last months in the nursery class. Again this was for

organisational reasons, and here the bias would be in the opposite direction to that cited earlier - also with minimal effects on the main conclusions.

All the sample children were given the post-test battery (reading, mathematics and Piagetian assessment) in late May, June and early July. These tests, as with the earlier tests, were first administered at the pilot school to ensure the effectiveness of the instruments and the consistency of administration by E.

A detailed examination of the results of all the tests is presented in section 6.40. A few points about the sample and the testing need some discussion here.

The problem of attrition was not unexpected, and in the early part of the research project the losses were reasonable, approximately 90 per cent of the original sample surviving by mid-1977 and 85 per cent by early 1978. Some losses occurred at the outset, when the children of a small number of the interviewed parents could not be tested, for various reasons (irregular attendance, foreign language problems, excessive shyness, etc.), or alternatively in a few cases it proved impossible to find the parents at home for an interview, although their children had been tested. In the case of parents who left after attending programme meetings there was usually warning of departure; other losses were discovered when planning a further testing session. The reasons for leaving and the characteristics of those who left varied widely.

Three parents left the advantaged school during the main period of the research, compared with 27 parents who left the five disadvantaged schools. This small differential bias was confirmed in the analysis of the differences of the means of the attrition and remaining samples, on ten important variables. However, only in the case of one variable, parent reading behaviour, does the difference attain a one-tailed probability of 0.05. (On the basis of reports from other studies that it is usually the more disadvantaged who are the more mobile, the use of the one-tailed criterion was considered appropriate.)

Table 18 gives the analysis of the differences between the attrition sample and the main sample surviving at the time of the mid-tests. The trend towards a lower performance level on these variables is present but it is not strong and in several cases the differences are in favour of the attrition group. An alternative comparison is offered in the same table, where the data on the attrition sample are compared with data from the 129 parents remaining in the disadvantaged sample at the time of the post-tests. (This latter group excludes the single non-working programme refuser among the disadvantaged.) The second comparison shows smaller differences than with the previous comparison; however even these figures indicate a moderate trend towards lower mean performances among the attrition parents.



It is noteworthy that the biggest difference, in both sets of comparisons, occurs in parent reading behaviour, suggesting that it is not so much the child's levels of cognitive functioning or early attainment which distinguish between the attrition and surviving samples as the level of reading 'environment' surrounding the child in the home.

Table 18. Mean data for attrition sample compared with  
(a) main surviving sample; (b) main surviving disadvantaged sam.

	Attrition sam. n = 30 *	Main surviving sample n = 164 **			Main surviv. disadvd. sam. n = 129		
			diff.	t-stat		diff.	t-stat
E.P.V.T. s.d.	18.03 (10.83)	19.42 (9.59)	+1.39	+0.71 <u>p 0.24</u>	18.60 (9.30)	+0.57	+0.29 <u>p 0.56</u>
Maths s.d.	10.37 (4.23)	10.27 (3.78)	-0.09	-0.13 <u>p 0.55</u>	10.02 (3.17)	-0.35	-0.51 <u>p 0.70</u>
I.R.T. s.d.	4.57 (5.03)	5.99 (6.23)	+1.43	+1.19 <u>p 0.12</u>	5.43 (5.66)	+0.87	+0.77 <u>p 0.22</u>
W.P.P.S.I.							
Info. s.d.	8.13 (3.69)	8.63 (3.73)	+0.50	+0.68 <u>p 0.25</u>	8.24 (3.64)	+0.11	+0.15 <u>p 0.45</u>
Sents. s.d.	9.17 (3.67)	8.52 (3.76)	-0.65	-0.87 <u>p 0.81</u>	8.40 (3.87)	-0.76	-0.98 <u>p 0.84</u>
Pic.C. s.d.	9.50 (3.66)	9.93 (4.50)	+0.43	+0.50 <u>p 0.31</u>	9.78 (4.48)	+0.28	+0.32 <u>p 0.37</u>
Bl.Des. s.d.	6.47 (3.78)	6.20 (3.56)	-0.27	-0.38 <u>p 0.65</u>	6.08 (3.50)	-0.38	-0.53 <u>p 0.71</u>
Rdg.Behav. s.d.	6.69 (3.61)	7.84 (3.50)	+1.15	+1.64 <u>p 0.05</u>	7.77 (3.47)	+1.08	+1.43 <u>p 0.08</u>
Lang. Env. s.d.	5.24 (1.94)	5.53 (2.00)	+0.20	+0.73 <u>p 0.23</u>	5.61 (1.96)	+1.96	+0.85 <u>p 0.20</u>
Mat.Beh. s.d.	3.92 (2.30)	3.60 (2.16)	-0.32	-0.74 <u>p 0.77</u>	3.43 (2.02)	-0.50	-1.13 <u>p 0.87</u>

\* For Reading Behaviour and Maths Behaviour the attrition sample n = 26, and for Language Environment the attrition n = 25

\*\* For Reading Behaviour, Language Environment and Maths Behaviour, the comparisons are based on the final surviving sample of 159

Another way in which the attrition data was examined was to cumulate the scores of the lowest and highest groups within the attrition and main samples, for each of the first seven variables in the above table. The lowest and highest thirds of each sample were first compared; there appeared to be a slight trend towards lower scores in the bottom third of the attrition sample, and accordingly the lowest and highest sixths of each sample were examined separately. A particular purpose of this analysis was to examine whether there were statistical grounds for the impression that those who left the area included an undue proportion of both bright and slow children. The results do not indicate any marked grouping of bright and slow extremes within the attrition sample.

Verbal reports from parents and teachers presented a wide variety of reasons why 15 per cent of the original sample left the area in the first 16 months of the project. In some cases it was upward mobile parents who said they were moving to a "better area" - on occasions, derogatory racial arguments were also offered; in other cases parents moved because the housing authority had offered better accommodation elsewhere - or housing was available from some other authority; and in yet other cases parents left after suffering some marital or other social crisis which necessitated immediate departure.

The composition of the attrition group was also examined to find out whether there was any tendency for programme parents to leave the sample in greater numbers than non-programme (working) parents. While the proportion of programme accepters to working parents was approximately three to one in the original sample, the attrition loss of programme parents was less than twice that of working parents (18 to 12). There is thus no evidence of an undue loss of parents from the programme groups as a result of this phenomenon, although differential loss of experimental parents has been seen as a possible threat to validity in some other research studies. Anderson (1973) warns that failure to take account of dropout differences can prove a major source of error in reaching conclusions about the effects of educational treatments.

It has to be recognised, however, that the attrition sample as a whole, including a fair proportion of parents from all sample groups, has a somewhat lower performance level than either the remaining whole sample or the remaining disadvantaged sample, on two important variables - Infant Reading Test and parent reading behaviour. In only one comparison does the probability of these differences reach 0.05. With four other variables, E.P.V.T., WPPSI Information and Picture Completion, and home language environment, there is a consistent difference in favour of the surviving samples, the probabilities varying from 0.20 to 0.56. With the remaining four key variables, mathematics attainment and parent mathematics behaviour and WPPSI sentences and

block design, the differences are consistently in favour of the attrition sample, although generally small.

As explained above, there is no differentially higher attrition loss among programme parents. Overall it can thus be concluded that while there are differences in the attrition sample, especially in the field of reading and verbal variables, these are not large enough nor is their effect specific enough to any parent group to impair the research design and planned analysis of the data.

There was no evidence that the few losses occurring after March 1978 were radically different in character from the losses up to that time, and it was thus not considered necessary to carry out a further attrition sample analysis on the last 5 children who were lost up to the time of the post-tests.

Throughout the latter half of the research project considerable efforts were made to retain children within the sample, provided complete data were available on the nursery tests and parent interviews, and provided the children had remained at the sample schools until the end of the 1976-77 school year. Where possible children who left after that date were traced to their new schools, if the schools were within reasonable travelling distance from the metropolitan centre. The children were given the reception battery (mid-tests) at the new school, if these had not been administered prior to departure from the old school. All the traceable children were also given the final post-test battery at approximately the same time as the rest of the sample.

Of the final total of 159 children, 23 had been traced to their new schools and tested there, in some cases the schools being 100 miles distant from the sample area. Twenty such schools were visited and in all cases the fullest cooperation was given by the Local Education Authority and the school itself, for the testing arrangements. Children who had moved to new schools were given a data coding for later assessment of whether there had been any identifiable effects of the move on their school performance at this young age.

Since it was not possible to delay analysing the data until the end of the final tests - which could have extended the research period unduly - preliminary examination of the pattern of each set of variables was carried out on the basis of the sample surviving at the time when a particular data set had been coded. Thus the pattern of the Nursery battery was based on 176 children, the Parent interview pattern on 172 parents, the Mid-test pattern on 164 children, and all the remaining analyses on the final sample of 159 child-parent dyads. The correlational study, the path analyses and all the major conclusions are based only on the final sample. The data patterns are examined in section 6.40; the major analyses appear in sections 6.70 and 6.80.

One further point to note about the testing concerns the children's willingness and motivation in the test situation. It was decided to follow the practice adopted in many other programmes involving young children, namely to reward them for participation in the 'games'. As Sigel (1974) points out in a review of the difficulties of testing children, it is rather deceitful to try to motivate them simply to come and 'play games' when they know that they are not normally involved in playing games with relatively strange adults. The sample children were told that the rewards were for their hard work in these games. It was hoped that this would motivate them to the same degree that a teacher's urging might do in a classroom situation. The nursery children were each given a mini-packet of peanuts or sultanas; the reception children were given an inexpensive fibre pen; the post-test children were given a choice of inexpensive gifts.

In only a very few cases was there any difficulty in persuading the nursery children to come for the tests. The testing conditions and prior establishment of rapport have already been described in section 4.31. In each nursery class there were a few very shy young children; in these cases, 'leader' children who had already undergone the tests were asked to accompany their friends to the testing table, and the two children would sit there, often hand in hand, with the shy partner now willing to answer the questions and perform the tasks given. Out of the original sample there were three children who refused to be tested, and their right to this refusal was naturally respected.

The contributions made by the nursery teachers in easing the testing situation, in making the children aware of E's purpose and in helping with the more reticent children, cannot be overemphasised. The help was given generously and E's task made far easier than it would have been without their advice and full cooperation. The same comments apply to the reception class teachers; they too offered invaluable help and advice, although at that stage it was only rarely that shy children had to be persuaded to go for testing, or be accompanied by a bolder friend.

The testing situation itself was friendly and the formality of the protocol was minimised by the freedom of discussion.

A handful of vignettes, out of the wealth of possible quotes, suggest the flavour of this situation:

E, shuffling a set of Piagetian multiple classification cards, making small talk with the child: "Which do you like most at school, reading or numbers?"  
Child: "Dinner."

There was the child of limited ability who had hardly started reading, but

proudly identified 'flag' and 'castle' in the Southgate test; the Royal Jubilee had been a classroom project earlier in the year.

There was the extremely restless child who played and squirmed throughout the testing, and whistled in mock astonishment as each new activity or item was produced by E.

E, administering the Infant Reading Test, shows the flash word STOP and provides the protocol clue: "This is a word you see in the street."

Child: "Be careful!"

E, administering the Southgate test, points to the word we and waits for an answer. Child: "That's a rude word, ain't it?"

E, in the Infant Reading Test, points to Coca Cola and gives the protocol clue: "This is the name of something you drink." Child, a diminutive girl from a deprived home: "Champagne." Other children's offerings on this item included beer, vodka, whisky and tea, apart from a fair number who recognised "Coke".

One of the flash words in the same test read bubble gum. A bright 5½-year-old: "One shouldn't have 'bubble gum' there as you've already got 'sweets' up there (referring to the previous word), because bubble gum is a kind of sweet."

The search strategy of a child trying to puzzle out which was the right word (out of five indicated choices) in a Southgate item: "Wheels? It must be the last word because it can't be the second word, that would be too easy."

## 6.20 Parent interviews

Parents generally gave a warm welcome to E when he called on them for an interview. The letter sent to parents in advance (appendix B4) has already been described. In most cases parents had also been informed by their children of the 'games' being played with a newcomer to the nursery and of the rewards given to all those who played the prescribed games. In some cases the time of calling was inappropriate and an arrangement was made for E to call later. Interviews lasted from 15 to 75 minutes, depending on the interest of the parent(s) in the interview themes put forward by E.

## 6.21 Extraneous issues

There was a tremendous desire to discuss general school matters and many parents thought, despite the clear wording of the letter, that E was 'a man from the school'. Often matters concerning older children, or children at other schools, were discussed with E. It became evident that many parents, particularly among the disadvantaged, had a need to discuss these matters with someone who was seen as somehow representing or liasing with 'authority' in the educational field. In some cases tricky issues were raised and E endeavoured to present what might be termed the teachers' or school's point of view. Children's behavioural problems figured prominently, together with the teachers' responses. In all but two cases, which were referred to the school authorities concerned, the fact that these issues had been aired appeared to satisfy the parents; they were not seeking action or redress of complaints so much as the opportunity to talk about school issues with someone who was reasonably knowledgeable in that area. At the conclusion of the interviews more than half the parents expressed the wish that they could meet someone from the school in their homes at regular intervals - perhaps once a year. They saw such meetings, on their home ground, as a way of improving relationships and understanding on both sides. Many parents did not feel that Open Days provided the opportunity, privacy or encouragement to discuss problems in the way that a home meeting could do.

More personal concerns about the children were also raised by a number of parents. Mothers were particularly worried about what they termed the excessive activity displayed by their children; some children were described as being incapable of even sitting in front of the television for more than a few minutes at a time. Various parents claimed that their children were driving them 'mad' with the endless movement and refusal to sit calmly. A few reported that

their children were attending clinics to get help with language or behavioural problems. The reaction of parents to behavioural problems varied from passive acceptance - one mother was interviewed by E while an endless succession of her four young boys raced between and around the interviewer and the interviewed - to harsh measures to cope with what was sometimes seen as 'wickedness'; several parents boasted that they hardly ever bought toys for the children, because "the child just destroys them".

The discussion of matters only distantly related to the research theme had of necessity to be limited - except in a few cases of extreme concern. In general no more than one-third of the interview time was spent on extraneous matters raised by the parents themselves.

#### 6.22 Group allocation

The number of parents visited, the main groupings and the broad characteristics of these groups have been discussed in section 4.20. The characteristics of the attrition sample have been analysed in some detail in section 6.10. A fuller discussion of the parent groups and the parent-child dyads appears in section 6.40.

The allocation of parents into research groups was based on the procedures outlined in sub-section 4.226. In only three cases were parents allocated to a 'refusers' group on the basis of their response to the invitation to participate in a programme (working parents were in a category of their own). One of these three parents expressed hostility to the whole concept of parental involvement in the early education of the child; in another case a highly advantaged parent pointed out that she was sure the school would do a good job and as her older children had done well at school she did not think there was any call for her to take an interest in academic matters; a third parent claimed that she had lots of domestic duties and did not see the relevance of the proposed programmes. On the other hand, a parent who expressed interest and support for the idea that mothers had a part to play in preparing the children for school, but who was extremely busy minding four or five children, was allocated to the working parents group when it became clear that she would be unable to attend any programme meetings. Several other mothers, each with only one minded child apart from their own nursery youngster, attended programme meetings and were categorised in the normal way as programme parents.

In several cases, all involving ethnic minority parents, the fathers refused to allow their wives to attend programme meetings, since a man (E) would be organising and attending the meetings. A few of these fathers said

they would attend themselves, although in fact none of them did so. On the other hand there were several cases where fathers attended because their wives were working, and in two cases father and mother took turns in attending. While cultural mores kept several Asian mothers away from the programme meetings, other Asian mothers who attended meetings were among the most reliable and conscientious participants.

#### 6.23 Goals and ethical issues

When conducting interviews the guidance offered by Newson and Newson (1976a) in a discussion of their own experiences offered some pertinent cautions about interpretation of the behaviours and attitudes of disadvantaged parents. There was however a somewhat different emphasis in the present research. Whereas the Newsons sought to gather interview data for an interpretation of the parental roles within particular social contexts, this research sought to gather interview data on parental behaviours and attitudes specifically in relation to the child's early academic attainment. The goal was to quantify these characteristics as clearly as possible in what is a somewhat loosely defined area, regardless of whether there were valid reasons why parents faced with major social handicaps could not practice what are often seen as important and desirable behaviours, such as reading regularly to toddlers and young children.

Interpretation of the failure to provide a sophisticated verbal and intellectually stimulating environment is an area that merits much further study, beyond what has already been offered by workers such as Bühler (1943), Douglas (1964), Vernon (1969) and White (1977). The present study aimed to quantify the relevant home variables, as far as possible, and to relate them to final child outcome as well as to the level of participation in the parent programmes, within an analytical model which takes full account of the child's characteristics.

Although the interpretation and justification of what might be seen as inappropriate developmental behaviours (by parents) were not major issues in the present study, the ethical problem raised its head in another form. There were parents who reported various behaviours and ideas contrary to the researcher's hypotheses - for example, the view of many that the parents' only responsibility to the children was to feed and clothe them and make sure they got as far as the school door. With programme parents this attitude could be discussed at group meetings; but with working parents who expressed such views there was the temptation for E to demur politely and suggest that perhaps talking and reading to the child might help in the child's development and prospects at school. While it was considered justifiable to take a neutral



stand in regard to such views - since the children of such working parents would serve as an important comparison group - it was not always easy to escape responsibility in discussions of these matters. One couple, both working, boasted that they were already teaching the child "words and pictures". The mother displayed picture captions written in minute letters. E felt constrained to stress the need to write in large and lower case letters. Another father assured E that he gave his son strong beatings if he didn't sit patiently listening when the father read (long) extracts from the Bible every evening; again E attempted to discuss the traumatic effect that such beatings might have. Later, after discussion with the nursery teacher, the matter was referred to a sympathetic social worker since the child was already showing disturbed behaviour at school. These were typical of incidents in which ethical priorities served to 'contaminate' the purity of the different parent groups. Such problems inevitably arise in action research situations.

Other sensitive situations arose in regard to some of the sample fathers. There were unemployed men sitting at home, sometimes self-consciously, while the interview was conducted with the mothers. There were fathers who disassociated themselves from any responsibility for the upbringing of the children. In a few cases fathers sat glued to the television set, at high volume, while the mother endeavoured to answer E's questions. But in the majority of cases the father, when present during the interview, took considerable interest and offered valuable insights on the issues discussed. The father's apparent educational level - based on the level of discussion - did not appear to be a strong determinant of interest or disinterest, although educational level did appear to relate to the nature of the interest taken in the child.

There were clear cultural imperatives in the interviews with parents of minority group parents, and E took care to recognise limitations on the freedom to interview - particularly in the case of Asian mothers. The main ethnic minority consisted of parents from the West Indies, or parents who had been born in this country of West Indian parentage. Many of these people, and also parents stemming from African countries, regretted the absence of grandparents within the family environment in Britain. They felt that this absence had a detrimental effect on their pattern of upbringing, although in only a few cases did parents see this as relating to developmental issues such as language. While there were indications of Black consciousness in the religious and political spheres - from snippets of conversation - it was only in the case of a few middle-class Black parents that there was any evident cultural awareness in regard to reading matter bought for their children. These parents regretted that bookshops carried so few reading books featuring Black children, although it was pointed out by E that all the local libraries carried many such books.

6.24 Comments on nursery education

The interviews with the 172 parents in the sample at the time of coding offered valuable insights into what such parents think about nursery education. Projects such as that of Tizard, B. et al (1981) describe the work done inter alia to deepen parents' awareness and understanding of the goals of nursery education. The interview protocol (Appendices B) contained questions designed to elicit views on these goals.

In discussion of the nursery class issues it was noted that many parents - the majority in some of the more disadvantaged areas - continually spoke of their children's attendance at 'playschool'.

The parents' answers to the main question about the goals of nursery education ("In what ways do you think nursery class has been of (much/some/some little) help?" - the relevant adjective being based on the parents' previous answer) were coded in terms of one or more categories - custodial care, social care, language, general development, educational and other goals.

Only a handful of parents proffered custodial care as one of the goals - some 6 per cent. A typical answer here was "nursery class gets the children away from under the mother's feet".

Close on 60 per cent of the sample offered socialisation or other forms of social caring as their impression of one purpose of nursery education. Two-thirds of the advantaged school's parents offered this reason, and three-quarters of the parents at the least disadvantaged of the five disadvantaged schools.

Just over 27 per cent of the sample offered language as one of the goals of nursery education. The highest percentages came from two schools where the nursery teachers were particularly aware of the developmental goals of the nursery class, although in other ways these two teachers stood at opposite ends of the philosophical spectrum on open or structured education.

Nearly half the sample (46 per cent) spoke of the child's general development in one or other form. Again it was noteworthy that two of the three highest percentages came from the same two schools cited above, suggesting that when a nursery teacher is particularly conscious of the developmental function of the nursery class, this may communicate itself to more parents than would otherwise be the case. However, the increases in percentages were not large enough to merit anything more than these speculative comments.

Only 13 per cent of the sample saw the nursery function as educational. The highest percentage (and even this was only 21 per cent of the parent sample at that nursery class) was found at a school where number games and some other

broadly educational activities had been offered to interested children.

The socialisation function of the nursery class was thus clearly uppermost for many of the parents. The value of learning to mix with other children of the same age was often mentioned - not only by advantaged parents but also by those disadvantaged parents living in high rise flats or other relatively isolated environments. A number of parents mentioned social goals such as teaching the child 'manners' or "discipline to get them into line for school"; "nursery class has calmed down my child"; "it gives \_\_\_\_ much help in his social development and stability".

Parents who spoke of the developmental function of the nursery class often mentioned the fostering of creative activities - "there's not much discipline or general development in the nursery class, but it's very creative". Others said that their child had become more artistic; one felt that although the children played games too much of the time, they did learn handicraft skills; another pointed out that her child was very bored at home before starting at nursery, but was now more interested in her surroundings because of what went on in the nursery environment.

In contrast to the appreciative comments in regard to other nursery goals, there was some criticism of what parents considered to be the failure of the nursery to provide an educational experience for the children. Some stated that nursery was useful as a preparation for 'big school', but more expressed doubts: "nursery class doesn't give a broader educational preparation, but it has helped \_\_\_\_ to find his own identity"; "I wish nursery class would do more for the child's education, like reading and writing"; "a lot of the child's interest (in educational activities) may be gone by the time he enters school" (this parent played educational 'games' at home with her child); "all they do there is play - it would be better if the children were taught their numbers and how to tell the time". The criticism was occasionally indiscriminate; one parent complained that "they just play with toys and games at nursery class", this interview being conducted under the difficult circumstances of a television set playing at full volume throughout the interview. Several parents ventured the opinion that the reason why nursery classes were so non-educational was that this was laid down by the educational authorities. Only a few parents spoke of the wider educational function of the nursery class as providing a period of general preparation for the more specific academic work in the reception class.

Perhaps the most original comment came from a mother who said that she and some of her colleagues felt that the recently introduced extra year of compulsory schooling should have been added at the beginning point, with four-year-olds, rather than at the other end of the school years. There should be

more freedom of choice for the older age group and correspondingly more emphasis on the potential of the younger age group (the four-year-olds) at a time when the children's levels of curiosity, awareness and the ability to absorb information and experience were so high.

## 6.25 Reading activities

One of the key features of the research project was the pre-reading programme to be given to certain groups of the parents. Accordingly questions on parent reading behaviours and attitudes played an important part in the interview protocol.

The analysis of the data on each parent's reading behaviour in relation to her (or his) child forms an important part of the statistical sections and path analyses later in this chapter, and will thus not be dealt with here. The comments provided many insights into parents' individual behaviours and their attitudes about responsibility for the initial stages of the child's learning to read.

A key question, which was approached indirectly so as not to yield an 'acceptable' answer, centred on whether the child was read to, and if so, how often. The frequency distribution and the various correlations of the combined score for these questions showed its reasonable validity in relation to the reading attainment of the child. As expected, there was a sharp difference between advantaged and disadvantaged homes in the amount of reading and the quality of the books used for reading to the child. At one extreme were a few parents who read to the child interesting excerpts from any source - even from quality papers such as The Times; at the other extreme was the mother who assured E that she read a great deal - to herself - but never read at all to the child "because the child is not interested in reading".

The books used for reading to the child were extremely limited in variety among the disadvantaged sample. Ladybird was mentioned more than any other type of book; reading books inherited from an older child, or even from the parents own childhood days, were also quoted; Bible stories and the Koran were among the other sources. With some parents who mentioned having "read books" to their children, or "bought books" for them, it turned out that these were simply picture books. Books of nursery rhymes were mentioned quite frequently. Modern children's books such as the Dr. Seuss series were only occasionally cited. One father said that he read the Bible and Enid Blyton's 'happy stories' to his four-year-old son every evening; the boy, who was present, interjected "it's boring". This boy was not allowed to play with

any other child outside school hours. Several parents explained that they did not buy books for their children any more because the youngsters had "torn them up".

Older siblings played an invaluable reading role in a considerable minority of families, particularly in cases where the parents had a limited degree of literacy. The warmth of sibling relationships and the role of an older sister in reading to younger children was particularly evident in West Indian families. In a few families it was an outside relative - a grandmother or aunt - who read to the target child. One of the most moving cases was that of a four-year-old (nursery child) who 'read to' her 10-year-old brother every day - making up imaginary stories from a picture book. The boy was mentally handicapped and the little girl had boldly taken on responsibility for developing the mind of a boy twice her size.

Two questions aimed at discovering whether parents considered that it was their own or the school's responsibility to start the child on its initial stages into reading yielded many interesting answers. It was evidently a highly debatable issue and the later analyses of this combined item, scored as Parent Reading Attitude, did not show any strong relationship with level of advantage or other parent characteristics or child outcomes. Despite the lack of validation in terms of these other variables, the questions brought out a wide range of perceptive comments.

There were parents who saw the task of teaching reading, even at the most basic level, as entirely the responsibility of the school; some even thought that it was the school alone that should read stories to the child. At a more sophisticated level was the father, himself a former teacher, who did not read to his children because that was the school's function - whereas the teaching of numbers was a father's task. Other parents averred that they had enough work of their own to do; their only responsibility was to get the child to school.

Among the contrary views was the argument that the parents were with the child more than the teacher was and they were therefore better able to help the child to start reading. A parent emphasised the pedagogical value of the one to one parent-child situation. Several parents thought that the nursery teachers should offer the parents guidelines on how to teach their children to read, since parents had more time than the teachers to do this work. One mother said that parents should help children when they were ready to learn to read; if such work was left entirely to parents there would be many backward children; on the other hand the schools were inclined to say 'leave the education to us', but didn't always ensure that that occurred. A few parents reported traumatic experiences - in only one case was this a first-hand

report, from a non-sample school - of teachers angrily denouncing parents who had brought children to infant school already able to read.

The question of the right method appeared to many parents as the biggest obstacle to their helping the child. They did not know what method the school preferred. One father, a loyal supporter of his school, said ruefully that each of his three children had been taught by a different method - he cited 'Janet and John', 'Ladybird' and 'Breakthrough'. He didn't know how to help his fourth child. Parents claimed that the schools taught reading 'differently' from when they were young.

The possibility that the child might have to learn to read using i.t.a. was mentioned by a fair number of parents, although none of the sample schools used this method. (Two of the other 20 schools to which sample children moved did however use i.t.a.) In some cases E had to assure worried parents that there was no evidence that children suffered from learning to read with i.t.a.

Parents who did some preliminary reading work with their children spoke of a variety of methods. Flash cards were particularly popular. A number of children were taught to write their own names (sometimes quite informally by older siblings). However, such parents were a small minority in relation to the sample as a whole. Some parents expressed the conviction that it was their responsibility to start their children on the road to reading, but recognised their own educational limitations, such as a minimal level of literacy. One impoverished mother, interviewed in the family's sole living/bed/dining room, with a large bed occupying three quarters of the floor space and a television sited at the foot of the bed, regretted her inability to help: "Four small children, and just me alone".

An unexpected finding was that a fair minority of children had watched the adult literacy television series "On the move" and found it a highly enjoyable experience, widening their interest in and awareness of reading. Parents of these children were emphatic that the programme had been of great interest to the children.

Questions relating to the use of the various public libraries scattered around the sample area yielded a high level of seeming misunderstanding, and suggested a very low level of library usage by the disadvantaged parents in the sample. Parents were asked whether they belonged to any libraries or whether they got books from libraries. Even positive answers were not always evidence of library usage; probing on the basis of the original answers showed that some parents gave a positive answer when in fact the 'library books' were simply books brought home from school by an older child, without active intervention by the parent.

Some of the problems in regard to library usage in the area have been dealt with previously in sub-section 4.223. Even those disadvantaged parents who did make use of the nearest public library were unenthusiastic in their descriptions; their overriding concern was the safety of the books and the danger of damage. The usefulness or interest of the books was seldom mentioned, other than in a few cases where an older sibling was an active and regular library user. Even here one parent reported that an older daughter had been so fearful of returning books late that she had stopped using the library.

E made visits to the children's librarians and higher level executive staff at several of the larger public libraries in the area and was assured that parents were given every encouragement to take out books; there was little or no criticism if a children's book was brought back late or if it had a few pencil marks or other minor damage. The librarians found it hard to accept that there was any justification for the disadvantaged parents' expressed fears of harsh library rules and administration. The different viewpoints on this issue could not be resolved and attempts to develop parental use of the libraries during the group programme had only limited success, other than among the advantaged parents who had in any case been making use of their three nearby libraries prior to the start of the group meetings.

#### 6.26 Mathematics activities

Obtaining information about the parents' mathematical behaviour in relation to the children presented considerable difficulty. For many parents the concept of number-related activities was a strange one, except for the basic task of teaching a child to count. They could understand the concept of teaching a child to read or offering it reading-related activities with words and letters, whether or not such activities were practised. But the playing of number games was seldom reported among the disadvantaged. With many parents the discussion about such activities yielded either no information or at most a report on the chanting by the mother of the number of steps as mother and child wearily climbed one or more stairways to reach home after a shopping expedition. Some parents cited the child's set of Lego blocks as evidence of the home's interest in number activities. Judging by the high proportion of parents who said that their children played with Lego blocks (in later questioning about toy purchases) the use of these blocks is nearly universal among sample families; the extent to which they develop spatial awareness in the child is a matter that might merit research beyond the confines of the toy industry.

Similarly to what occurred during discussion of the possibility that parents might initiate the child's first steps into reading, a small number of mothers suggested that nursery teachers or the schools themselves might show parents how to teach 'numbers' to the child.

The difficulty of eliciting a more wide-ranging score of mathematics behaviour on the basis of interview questions was not resolved in the present research. The unexpected finding that the mathematics programme had its only real effect (for the disadvantaged sample) in the area of mathematical and Piagetian concepts suggests avenues for further investigation, although judging by the problems experienced in making parents aware of the conceptual dimensions in the mathematics programme, eliciting of parental reports on such behaviours may prove even more intractable.

#### 6.27 Attitude to school and school work

The quantification of the parents' attitude towards the school and of the parental view of where the main responsibility lay for inculcating a sense of willing response to the work demands of the school environment was even more difficult to realise than was the quantification of the parents' attitudes towards the development of initial reading skills in the child.

This was partly due to the known difficulty of scoring attitudes and the limited relationship of these attitudes to actual behaviours. But it was also because the wide range of views expressed appeared to bear little relationship to the child's school performance; the views were rather a reflection of personal philosophical stances in which issues of 'discipline' and the 'structure versus freedom' controversy came to the fore.

Had the target child already been in the main school it would have been possible to ask behaviour-related questions such as whether the parent brought the child to school herself, how often she talked to the teacher and whether she played any part in school activities. In the nursery situation such questions would not have yielded valid answers reflecting parental school attitudes or voluntary behaviours, since it was expected that parents should bring their own children to nursery (and in at least one class insisted upon, unless approved alternative arrangements were made). Nursery teachers and assistants also went out of their way, in all the classes, to relate to parents at the start and end of the nursery sessions.

Another problem in quantifying school attitudes was that two of the three items related to parents' views of the nursery class and its function; parents who thought the nursery curriculum should be more closely aligned with the



school's academic goals scored low on these two items even though such parents might be very supportive at the school level.

Despite these limitations, responses to the items, particularly to the question as to whose responsibility it was to prepare nursery children for the school's demands, offered interesting insights into parents' views on these matters. A great many parents considered that the school had the main responsibility; "schools know so much more than parents about discipline, work and so on"; "preparation for the work situation in school is the school's responsibility - home shouldn't be like school"; "getting the child to learn to work is the teacher's area - I have passed on the child to the school"; "parents have no time to supervise their children's work". A few parents argued that children still 'played' a great deal in infant school and only really got down to much learning in junior school. A smaller number of parents saw responsibility for the preparation of the child as the joint task of school and parents.

The evidence of Hewison (1978) suggests that parents' behaviours in regard to their share in or responsibility for their children's school work becomes more crystallised and can be assessed in the primary years (particularly in the degree to which they listen to their children's reading at home each day); such behavioural scores - rather than parental attitudes towards their children's academic progress - serve as one of the predictors of reading performance.

Discussion of responsibility for preparation often moved into the area of discipline, where views were more clearcut and generally in favour of stronger disciplinary control over the children. Such views were expressed regardless of whether the particular children's behaviour at home suggested a sensitive degree of control, or the almost complete absence of checks on any behaviour that did not directly inconvenience the parents. It could be argued that whether or not parents had succeeded in inculcating self-discipline in their children, the schools were expected to succeed in engendering this virtue in almost any child sent there.

Attempts to elicit responses on the nature of the disciplinary control exercised at home were abandoned after it became clear, in a large number of the early interviews, that parents were presenting acceptable answers rather than the reality of home life. One incident highlighted this problem. A mother had been asked how she responded when the child refused to do what it was asked to do. Mother: "I give \_\_\_\_ an explanation and if it doesn't work then I switch off the television". Boy friend (present during the interview): "Yes but you also give her a smack and then hold her in your arms and explain".

At times parents voiced pessimistic comments on the problem of transition from the nursery environment to that of the school classroom. "Some children hate school compared with their (pleasant) experiences in the nursery class - it's an awful adjustment for them, especially those who like to play."

#### 6.28 Television viewing

Although the quantification of the child's television viewing time was based on E's subjective judgement - the 'dominance' of television in the home at the time of the interview, the way in which it was discussed, the programmes reportedly watched by the target child, and similar indicators - a comparison with parental estimates in a miniature analysis described in section 6.50 suggested reasonable reliability for the estimate.

It became apparent that the amount of television viewing bore only a minimal relationship to the child's cognitive levels. This may relate to the differing quality of the viewing process as experienced by different children, and that in turn could be based on the child's own cognitive and experiential input into the process. One of the brightest children in the sample was reported by her mother as being glued to the set for a large part of the child's free hours; the mother considered that the child's verbal precocity came from the great number of words she learned from TV programmes. Another mother of a bright child said that when there was nothing on television her child would sometimes sit staring at the myriad dots on the blank screen. On the other hand there was the impression that for some of the most backward children TV served mainly as a pacifier rather than an educational experience. It could be argued that their own verbal experience had been so limited, in terms of interaction with parents and siblings, that the television viewing could make but little contribution to the expansion of such children's linguistic or cognitive capacity.

On the assumption that E's estimates of television viewing were reasonably accurate - more sophisticated and time-consuming methods, such as fitting the child with a chest microphone to record a typical week's viewing, would be necessary for greater accuracy - the relatively low correlations between TV Time and the cognitive and academic criteria used in the study suggest that the effects of viewing on child performance are as much dependent on child input, in some as yet unquantifiable way, as they are a factor of the time spent viewing.

There are some unusually interesting findings in an American study (Perney et al, 1978) which corroborate several of the somewhat unexpected findings of the present study on the relationship between television viewing and academic

attainment, and suggest that despite the relative smallness of the correlations they have a certain validity and merit further examination in studies more specifically geared to the effects of television viewing on attainment. These issues are dealt with in more detail in section 6.80.

The estimate of the mean television viewing time for the sample was well over  $2\frac{1}{2}$  hours a day, with 31 per cent of these children watching more than three hours daily. (The variable used in the data analysis was scored negatively on the hypothesis that the amount of viewing would bear an inverse relation to academic performance.) The mean figure of 19.24 hours a week, for a 4-year-old sample, is a little lower than that of a recent British study (Durden-Smith, 1978) which found that two out of three children in the range 7 to 17 watched more than 21 hours a week, with an average of 22 hours for a sample of over 1,600 children. Another recent study (Ghikas, 1978) on a sample of 3,200 mothers whose children watched You and Me, a B.B.C. children's programme, noted that half the sample reported their children's viewing as less than two hours a day, with the remaining half viewing from 2 to 6 hours daily; only a very small number were reported as watching for more than 3 hours a day. A major American pre-school TV study (Lesser, 1977) quoted figures ranging from 22 to 24 hours for first-grade and pre-school children. The other American study already cited (Perney et al, *ibid*) reported total viewing hours of 17.2 for boys and 16.5 for girls; this study was based on a middle-class sample of 200 kindergarten children. It is interesting to note that the equivalent British middle class sample in the present research - the 29 children at the one advantaged school - showed a mean viewing time of 17.9 hours a week, which compares well with the American data.

Questions about the kinds of programme viewed revealed that most children watched the then current child favourites, such as "Bionic Man"; longer-running series all had their quotas of child watchers. Popular musical shows were also mentioned by many. Only in the case of the more erudite youngsters were adult programmes cited by the parents. The one exception to this was the adult literacy programme "On the Move" which, as explained earlier, had been eagerly watched by a large minority of children. Parents expressed surprise over this phenomenon, as none of them had suggested that their child should watch the programme and it was only by chance that any child became aware of it.

A considerable number of parents mentioned that they and their children had enjoyed watching Sesame Street, when a few extracts were broadcast by the B.B.C. or when Independent Television ran a more regular series for a while. All who had seen it expressed regret that it was not now a regular feature of the British viewing scene; they considered that children learned a great deal

from it, despite its American flavour. The most widely watched children's programme appeared to be Playschool and only rarely did any parent say that their child did not watch this; in most cases the viewing of Playschool was regular rather than sporadic.

A majority of parents saw television as a healthy medium, contributing in some way to the child's development. Some saw it as a kind of universal educator; "it teaches the child everything - she is so attentive while watching"; "it keeps them amused"; the value of TV was "in explaining words to children - the child would be lost if he had to rely on parents' reading to him so that he could learn new words".

A few of the advantaged homes had no television, as a matter of principle; such parents felt that the child had so much experience to enjoy in the young years that television would intrude upon this learning process. There were also a handful of parents, advantaged and disadvantaged, who saw TV as a harmful factor in the child's development; "it is harmful in the sense of community because it alienates the child from experience".

These few critics apart, television appeared to serve as one of the cornerstones of family life within the sample; a number of parents said that children (in general) took in more from watching television than they did from reading.

#### 6.29 An ideal parent?

It would be invidious to argue that any one kind of mother was an ideal parent in terms of fostering the child's development. Even the goals of development are not a matter of wide agreement. But if the goal of an early introduction to reading and number activities, in the form of games and interaction between mother and child, were to be seen as desirable, an ideal parent might offer a compound of the activities listed below - each of them reported by one or more mothers in the sample:

- a. Spends between one and two hours a day interacting with the target child.
- b. Tries to organise her working day so that domestic chores do not occupy all the limited time she has for her child; or alternatively ensures that husband or boy friend share in the chores as well as interacting with the child.
- c. When the child seems ready or interested, talks to it about words and points out a few meaningful names on household objects, or plays 'I spy' with it.
- d. Treats bus journeys and shopping expeditions as important outings for the

child, explaining scenes of interest, pointing to the occasional symbol, such as a bus number or destination or the name of a shop that is visited frequently.

e. Plays finger games with the child, developing number awareness.

f. Enables the child to play with water and common domestic objects which match or fit one another.

g. Helps the child to choose television programmes on the basis of the Radio Times and TV Times, or the daily newspaper's programme summaries.

h. Buys educational toys from shops such as the Mothercare chain - by far the most highly regarded - and introduces further reading and number activities as the child shows increased interest in the relevant fields.

j. Endeavours to foster interest in these and other educational areas rather than waiting for the child to take the whole initiative.

In the latter part of the research project one of the sample schools in a disadvantaged area made available to interested parents a booklet prepared by a number of teachers (Warlow, Alexeiev et al, 1977). This booklet offers a wealth of practical ideas to parents, including many of the points outlined above. But the problem remains - and this was recognised by the school in question - that parents who most need the lessons in that type of booklet are also the kind of parents who are least likely to ask for a copy, or if given the booklet, would be the least likely to read or understand the principles underlying the suggestions made there.

### 6.30 Parent programmes

A report on the attendance at programmes covering nearly 100 parents in six different schools, together with the realities, problems and parental views of these programmes, is not easily condensed into its essential or most relevant features. E's personal relationship with that number of parents, in 25 small groups, inevitably brought out a wide range of the features of life in the disadvantaged areas of an inner urban environment and also the sharp contrast with life in a relatively advantaged area nearby. During the nine month span covered by the interviews and group meetings there were many intruding events of much greater import to some participants than the programmes themselves - unemployment, marital break-up, house-moving, reduction of social security benefits for one or other reason, and similar crises - but also happy events, births, new jobs, street parties during Royal Jubilee celebrations, and the excitement surrounding school functions.

Parents were made aware that they were taking part in a research project and it was some surprise to them to find that the project covered more than their particular school. Nursery teachers played a major part in fostering parental cooperation with E, especially in the last week before the start of the programme meetings. In terms of attendance at meetings the most successful school was one where the nursery teacher had asked parents to meet her briefly at the start of the morning and afternoon sessions one day, and had spoken to the mothers about the purpose of the study, the planned interviews and the proposals for a 'parent programme' to enable parents to give more educational help to their children.

The advance preparation of the programme materials and the organisation of the meetings took up a great deal of time; the fact that identical reading and mathematics programmes were being offered at the different schools brought this task within the bounds of possibility and also ensured a reasonably similar 'experimental treatment' at each school.

### 6.31 'Selling' the programmes

From discussions with nursery teachers and the parents themselves it became evident that E would have to 'sell' the programmes to mothers who had a variety of alternative activities to draw them, ranging from the unexpected arrival of 'Nan' (the grandmother), a shopping expedition, a visit to a friend

or simply doing the chores at home and watching a little 'telly' until it was time to go out again and bring the child back from nursery class. The popularisation by E of the idea of attending meetings on how to aid the child's educational development was initiated mildly at the parent interviews and then followed up with the advance programme letters and activity materials sent to intending participants (section 4.33 and appendix C2). Posters were placed in all the nursery classes, although inevitably these had to contend with the many other posters already claiming the attention of parents.

On each day of meetings at a particular nursery additional posters were placed in a prominent position as well as the names of parents in the relevant groups and the times of particular meetings. However E placed a limit on the degree to which the programmes were to be 'sold' and there was no attempt to pressure parents into attending. As soon as parents started to arrive E would go into a far corner of the nursery and talk to the early child arrivals, so that group parents who did not intend to be present or who had forgotten the meeting and made other arrangements for the day were free to leave without the embarrassment of seeing E or explaining their inability to attend.

Even this cautionary policy could be questioned. Parents who were regular and committed participants confessed that it was very easy to forget the meeting day and to ignore the poster on the door if their minds were taken up with other matters. Some parents said that seeing E in a corner of the nursery was the first jolt to their memories about that day's meeting. The problem of remembering the meeting day one or two days in advance (so as to avoid making alternative arrangements) was one that dogged committed parents throughout the four months of the programmes. There was no way in which the committed, who would like to have been reminded more pertinently, and the less-committed, for whom attendance was a matter of a last minute decision if there was nothing more exciting, could be divided into groups for differing levels of reminder. Even the most committed parents had days when other obligations were so pressing that they were glad of the opportunity to miss the meetings.

Although one nursery teacher voluntarily went out of her way to remind parents a day in advance - contributing to the resultant high attendance previously referred to - in most cases the organisational problems of seeking out particular parents amid all the other problems that had to be dealt with as parents arrived or departed each day made this clearly inappropriate for the nursery staff. It was a particular aim of E not to burden the staff with additional work apart from the information requested from them, the distribution of initial letters and their approval of testing and meeting arrangements within the nursery environment.

The only other method used to improve attendance was the sending of letters to parents who had missed meetings. There were only two such letters (reproduced in appendix C2) and they were carefully worded so as not to pressure parents into attending. The first letter (appendix C2, 5a) was sent to parents who missed the initial meeting. Great stress had been laid on attending that meeting, for obvious reasons, and from the limited response it was apparent that most of those who missed this keynote meeting did not intend to participate in the programmes. The second letter (appendix C2, 6) was sent to parents who had attended one or more of the first few meetings, but missed either the third or fourth meetings. These letters had a good response (see next sub-section). No other letters were sent to parents apart from the notification to three groups of unavoidable changes in the dates of a meeting.

### 6.32 Attendance pattern

While the composition and characteristics of the programme and non-programme groups are discussed at some length in section 6.40, the broad divisions are presented here briefly.

Sample surviving at end of project:

	Total	Disadvantaged area schools	Advantaged area school
Parent-child dyads	<u>159</u>	<u>130</u>	<u>29</u>
These 159 dyads made up of:			
Programme accepters who attended min. 1 meeting	99	80	19
Programme accepters who attended no meetings	26	21	5
Working group parents	31	28	3
Programme refusers	3	1	2

The 99 attenders had been randomly allocated to the following groups:

Reading programme	60	48	12
Mathematics programme	39	32	7

These 99 attenders .... /



These 99 attenders could also be divided on the frequency schedule:

	Programmes total	Reading programme	Mathematics programme
2-weekly schedule	83	47	36
6-weekly schedule	16	13	3

---

The programme attendance patterns can be looked at in various ways. At an early stage, with a larger surviving sample, there were 140 parents who had accepted the invitation to attend the programmes. Of this total, 106 came to the first and/or subsequent meetings and 34 failed to attend any meeting. As explained earlier, in several cases these non-attenders were parents who were planning to leave the sample schools and did not consider it worth attending only one or two meetings.

Referring from now on only to the final surviving sample, there were 99 programme attenders and 26 who failed to attend. The attendance pattern at meetings gives an indication of the rate of dropout during the four months.

Attended first or second meeting (or both):	99
Attended third or fourth meeting (or both):	74
Attended fifth or sixth meeting (or both):	60
Attended seventh or eighth meeting (or both):	51

The purpose of presenting attendance figures in this form is that it was inevitable that parents would miss some meetings for reasons of pressing duties elsewhere rather than due to lack of interest. Parents who lost interest simply failed to turn up again, while interested parents returned to later meetings, occasionally after several absences. Thus the number of parents at any single meeting was lower than the total still actively involved. The final meeting was attended by 42 parents and there were two apologies from parents who were unable to be present.

Normal attendance at meetings was scored as 1. In cases where parents missed a meeting but asked to be given the details and materials before or after a subsequent meeting, E spent up to 15 minutes individually with any such parent and the relevant parent score for the missed meeting was recorded as  $\frac{1}{2}$ . (The scores were obviously used for research purposes and were not given to parents or nursery staff.) When the parent was absent the score for that meeting was zero. In view of the importance of the ideas put forward at the first programme meeting, attendance here was scored as 2. The presentation of the sets of basic programme cards was also scored as 2, since any con-

structive use of either the reading or mathematics cards would be likely to aid the child's development in that area, even if the parent attended no further meetings.

For the reading groups as a whole, mean attendance for the 60 parents was 4.78 out of a possible total of 8 meetings. The equivalent mean mathematics attendance was 3.84, pointing to the lower attraction of that programme.

In comparison with many other adult education programmes, the attendance of just over half the parents at one of the two final programme meetings of their groups can be regarded as reasonably satisfactory. The attendance pattern shows a gradual dropping off, rather than a major fall after the first meeting and a levelling out thereafter. Nevertheless the drop after the initial attendance was somewhat higher, as some parents found that the programmes did not interest them. This was more evident in the mathematics groups, where parents were less able to see the utility of the programmes for their children's development, or alternatively were less able to cope with the activities proposed, many of which may have seemed remote from what is normally regarded as 'number work' - despite the explanations given by E.

Out of the 25 groups which were set up at the six schools, ostensibly 22 were still in existence at the end of the meeting period. This does not give an accurate picture, however, since 10 of these groups consisted of a single mother (or father in one case). The remaining 12 groups consisted (at the end) of up to four parents. The initial intention to cancel meetings if any group was reduced to one or two members seemed inappropriate in the light of the keenness shown by parents who continued to attend. There is one minor but unavoidable distortion of the above figures. The attendances of the 16 parents in the six-weekly groups were scored in terms of the agendas dealt with and the related materials given to them at a particular meeting. This varied between two and three sets of agenda cum materials at the two-hour meetings - thus six-weekly parents scored two or three points for attendance at their longer meetings. There was no other way of scoring which could provide comparability between the two-weekly and six-weekly groups, since the groups covered the same programme ground.

It is also necessary to point out that the four mothers of twin or pairs of sample children were each recorded as two parents, though with identical home interview and meeting scores. This was a legitimate statistical device, since the alternative of equalising the scores of each pair of children would have been misleading.

The value of the reminder letters sent out to parents who missed either

the 3rd or 4th meetings, after having attended one or more of the previous meetings, was shown by the statistical count. Some 31 letters were sent out to parents who missed the 3rd meetings of their groups; 17 came to subsequent meetings, though there is little doubt that some would have done so even if no letter had been sent. A further 11 letters were sent out after absence from the 4th meeting, and 8 of those parents came to further meetings. The purpose of these letters was to reassure parents who might feel that one or two absences were enough to bar them from subsequent attendance, in view of what they had already missed.

### 6.33 The dropout phenomenon

The fact that a considerable number of parents dropped out of the programmes is worthy of some attention. The explanation that such parents were just not interested in their children's progress is too facile. There is always a small minority, at all levels of advantage or disadvantage, who may well have little interest in their children's progress. A great many more are interested but lack the ability or knowledge to turn this interest into effective behaviours on behalf of their children.

A considerable number of reasons were given by nursery staff, the parents themselves, or friends of those parents, as to why certain parents stopped attending meetings or even failed to attend any meetings. Social crises played some part, especially in the case of those with long-standing problems who attended no meetings. A small number took up employment (especially parents whose children moved into reception class before or at Easter 1977). One parent was transferred to different shift hours. Another parent was found to be developing cancer. There were at least four women in late pregnancy, the event being one of continuing interest both within their groups and among the nursery staff. (In four cases of extended illness or for other valid reasons, E visited otherwise regular attenders in their homes to give them the details and materials for a single meeting, or to cover the material from several meetings in one case.) The fact that some parents were planning to move house (and school) has already been cited. Others had to walk long distances to and from the school, often in inclement weather; when attendance at a meeting meant walking this distance four times in a morning or afternoon - if there was nowhere else to go when the meeting was over - it was understandable that a few of these parents gave up.

It would have been inappropriate for E or the nursery staff to ask parents why they had stopped attending, as some might offer any excuses rather than say openly that they found the programmes unsatisfactory. Perhaps half

of those who dropped out did so because of their views of the programmes. In one of the mathematics groups a mother arrived very late for the first session; when the demonstration of the idea of sets was shown to the group she protested that she was not a child; she attended no further meetings. Some of the mathematics parents in other groups asked if they could not rather be given a reading programme. One mother attended most of her mathematics group meetings but revealed to a colleague that she had kept the mathematics materials for the future while concentrating on developing the reading skills of one of her two sample children.

Some parents may have considered that the prospect of a long series of programme meetings (originally 14 were planned) was too daunting a commitment and opted out for that reason. Possibly the majority of those who dropped out for 'programme' rather than 'non-programme' reasons did so because for them the ideas may have been uninteresting or too difficult to comprehend; or alternatively too far ahead of their children's level. Only one or two may have dropped out because the ideas appeared to be too simple.

#### 6.34 Inherent problems

A basic rationale for the method underlying the present intervention project was to find out whether providing advice and guidance for parents of nursery children, in groups apart from their children, would enable the parents to apply the developmental ideas to the children in their own particular way, freed from the constraints of a method specifically demonstrated by a professional or para-professional in the presence of the child.

There are some problems about this rationale which are self-evident. When parents seldom if ever read at home themselves they are likely to have difficulty persuading their children to adopt reading practices enthusiastically. There were a few cases of marginal or concealed total illiteracy. A number of parents said openly at meetings that they were using the programme materials not for the target child but rather for older siblings who were making little progress at infant school; in a few cases the materials were used both for the target children and for still younger siblings who were advanced enough to take an interest.

There were problems at a deeper level which were harder to cope with. On a number of occasions it occurred that parents said at meetings that some activity was totally beyond their child's capacity; I would suggest trying again, perhaps in several months' time. Yet often such parents would report back two to four weeks later with the news that the child was now coping quite well with the activity in question. The difficulty in this situation is that

without personal knowledge of the child and home situation there is no evidence of whether the child is really far from ready for a particular level of activity, or whether it needs just a little prompting and repetition from the parent. It may not be far-fetched to suggest that in view of the child's close and still almost symbiotic relationship with its mother, a competent mother may have as deep an understanding of her child's learning capacity as any teacher does in the first years of schooling.

While it was not often that teachers in the sample schools expressed doubts about the proposed programmes, the one issue which was mentioned more than anything else was the danger of undue pressure by the parents. Of nearly 100 parents in the programme groups there were only two parents whose reports suggested any evidence of undue pressure. In the one case the parent was attempting to use the word cards in a totally didactic fashion, the child being asked to repeat the words said by the parent and to memorise the written words. It was easy to explain to this parent why such methods were counter-productive. Later the child appeared to be making good progress at school.

The only other case concerning a programme parent deserves special attention. It is the kind of case which is sometimes used by critics as a 'typical' example of what happens when parents are given ideas about the early education of their children, although in this case it was learnt that the child had been pressured long before the start of the programmes.

The mother, from an ethnic minority group, was devoted to helping a child whose cognitive levels had been found to be well below the sample average. This mother attended every meeting of her group but took no part in the discussions. One day she arrived at the end of the meeting and asked if she could stay on to be given details of what she had missed. Once alone with E, the story was told of an uncle obsessed with helping the child overcome his apparent backwardness. The mother was in a mathematics group. The uncle was using reading material that he had purchased earlier as well as number cards from the programme, giving the four-year-old boy up to half an hour's 'teaching' every day. The mother wondered if this was right; she said the boy seemed very stubborn and would not learn. For her culture group education was all-important and the uncle had been telling the child that he had to work if he wanted to get on at school.

E gave considerable attention to the mother's problem at this and two subsequent meetings. He explained that at most the child should be 'bribed' with the promise of a biscuit or other delicacy - rather than sombre threats about his future school career - to work for just five minutes a day with the mother, on simple number games or any other easy tasks. The uncle should perhaps be asked to leave the work to the mother for the moment.

The child moved into reception class several months after the end of the programme. Soon afterwards he was given the battery of reception tests. There was a considerable improvement in his performance and his performance on the attainment post-tests eight months later was quite reasonable. E congratulated the class teacher - a woman from the same ethnic group - on having achieved such success with a boy who had shown only limited performance in the previous year. The teacher denied that the main success was hers; she said that it had been due chiefly to the mother's efforts and her endless patience with the child that this major improvement had occurred.

In the sample of 159 there was one other apparent case of undue parent pressure. The mother expressed the intention of attending programme meetings, but never did so; she was convinced that she and her husband were doing everything necessary to ensure the child's rapid progress; she was also highly critical of the school's failure to push her child. There was little that E could do to alter the parent behaviours; only attendance at programme meetings might have enabled some influence to be brought on the mother.

The existence of a few per cent 'pushful' parents is often used as an argument why no parents should be permitted or encouraged to develop simple reading or mathematical skills in their pre-school children. There is reason to think that a pushful parent will exert undue pressure whether or not there is any formal encouragement. The value of guidance is that it is likely to check such behaviours, as in two of the cases cited, rather than to stimulate worse excesses. Guidance is also likely to encourage the great majority who think that they have no role in the early education of their children.

The most fundamental and intractable problem with a group programme is that by its nature it cannot be geared to the individual levels of a particular parent and child. This difficulty has already been referred to. There was no resolution of the problem during the project, other than the advice to parents to adapt the materials to their child's level of performance and interest in the activities. The only real alternative to this method was to offer individual guidance to parents; that was only possible in the case of groups which had dwindled to one or two participants. Parents were also informed that they were welcome to discuss problems individually with E before or after any of the meetings, but the time schedule and perhaps the self-consciousness of some parents meant that this alternative resource was not used regularly.

A particular difficulty arose in the case of some very immature children. Programme participants spoke about children who did not seem ready for any activity that required their attention for more than half a minute. Here the advice was for such parents to try out some of the easiest activities, such as colouring in pictures in the reader and playing (verbal) word games

or, for mathematics group children, to offer them some of the wide range of activities made possible by the equipment issued to parents.

One advantage of the group situation was that it provided the kind of communal environment where it was possible to discuss questions of future parent-teacher relations in a frank but sympathetic way. Some parents were very aware of the difficult situation faced by teachers with classes up to 30 infants; discussion by E focused on this and other problems in the classroom, such as the wide range of abilities and varying talents found in all children. The abiding importance of cooperation, even in the occasional situation where a parent and teacher might differ sharply, was mentioned on several occasions during the programmes.

#### 6.35 The programme meetings

A brief record was kept of each of more than 150 parent group meetings. As occurred during the parent interviews, there was often a desire to discuss problems concerning housing or other matters that affected their lives. Except in the case of urgent issues such as the stream of rainwater which had poured into a participant's flat the previous evening, a forthcoming tenants' rights meeting, or a petition for increased play-space for children on a housing estate, discussions of extraneous matters were limited to a maximum of five to ten minutes at any one meeting. However, full discussion was encouraged of all matters pertaining to child development and the parental role, whether or not these were closely related to points raised by E. Questions concerning child behaviour and socialisation came to the fore on many occasions.

Below follow selected extracts from the diary notes, expanded where necessary to clarify points of interest.

##### 1st reading meetings

Gp \_\_. Discussion of how to keep children's attention focused.

Gp \_\_. Confusion over names; unknown to E, one participant told a friend (a very recent newcomer to the nursery who had not been interviewed) that the friend's name was on the group list at the nursery door, so friend also turned up at meeting, took her named agenda sheet and sat there rather confused; the 'real' Mrs. \_\_ (identical name, but from totally different ethnic group) turned up late, was told by the (by now equally confused) E that she must be due at the mathematics meeting in half an hour's time; problem finally resolved and E later went to the home of the 'real' Mrs. \_\_ to apologise and give abbreviated meeting there.

Gp \_\_. Nursery staff wonderfully helpful; brought in tea; took over toddler

from mother, saying "Why shouldn't these young children be welcomed into a nursery class?"

Gp \_\_. One mother left meeting very early; clearly uncomfortable; said she 'might come' next time.

Gp \_\_. Useful discussion of ethnic language issues.

Gp \_\_. Found that two mothers had been trying to teach 'words' to their children; explained that it was too early for such work; alternatives suggested.

Gp \_\_. Three-year-old child threw temper tantrum during an otherwise highly successful (6-weekly) meeting; continued screaming for 20 minutes despite attempts of mother, other mothers and E to calm child, offer toys or biscuits; secretary from nearby office rightly came in to complain.

#### 1st mathematics meetings

Gp \_\_. Children saw their mothers (meeting in entrance foyer) through the glass door of nursery class; much crying from one child; decided to hold future meetings in medical room.

Gp \_\_. Successful discussion despite interruption from endless flow of children running through the foyer (and the meeting parents); ostensibly they were just on their way through but repeated forays by some children suggested it was the journey rather than the destination which offered the excitement.

#### 2nd reading meetings

Gp \_\_. Gap between mothers with 'successful' children and those with 'troubled' children; E tried to blur the distinction.

Gp \_\_. Sense of parental camaraderie despite interruptions.

#### 2nd mathematics meetings

Gp \_\_. Highly productive discussion between parents; E didn't want to intervene.

#### 3rd reading meetings

Gp \_\_. One mother very willing but has simple understanding and poor relationship with her child in general; needs lot of guidance on relating to child.

Gp \_\_. Father complained that parts of the programme belonged to infant school and said he would only offer his child what the child was ready to do; E agreed whole-heartedly with father's caution; father still unhappy.

Gp \_\_. Enthusiastic father boasted of how he had 'demolished' the many spelling mistakes in an essay by his junior school son; E expressed friendly



horror and explained that 'gently does it'; help the son only with the worst errors on any one occasion, and find things to praise in the essay; father conceded the point.

All groups. High level of discussion on parent's educational role and on society's attitude towards parents.

#### 3rd mathematics meetings

Gp \_\_. Mathematics ideas seem to be getting across well.

Gp \_\_. Mother worried that younger (target) child doesn't have promise of older child; E discusses individual differences and 'second child' syndrome.

#### 4th reading meetings

Gp \_\_. Problem with simple but willing mother; child has been assessed by E as well below average for age on cognitive skills; E lays down firm guidelines for parent to follow, to avoid misunderstanding; E considers that despite difficulties the child can only benefit by mother's interest.

Gp \_\_. Mother discusses frankly (the other participant was late) her child's hyperactivity; was upset that clinic had told her to give more attention to the child; says she is always occupied with the child, and totally exhausted by it; the child has been assessed by E as extremely bright; mother is a conscientious participant, with no evidence of lack of interest or attention.

#### 4th mathematics meetings

Gp \_\_. One mother objected to the sets and other mathematical activities; wanted more straightforward number work; her school is a hard-working but formal one with early emphasis on number skills; the issues discussed.

#### 5th reading meetings

Gp \_\_. Both mothers present had missed the 4th meeting but responded to E's reminder letters.

Gp \_\_. Great difference in levels of performance of the children of two mothers; mother of child who is doing well helped to highlight the achievements of the other child; second mother reassured.

#### 5th mathematics meetings

Gp \_\_. One mother arrived very late; expressed deep concern about her relative's heavy pressure on her child. (Incident described in previous section 6.34.)

6th reading meetings

Gp \_\_. West Indian mother said that her child, who had now started at infant school, was reaping 'marvellous benefits' from the programme.

Gp \_\_. Parents at this and other groups have apparently done little over the two-week holiday break; they said that the six-week holidays are better for such activities because children tend to get bored and are then more ready to try out the activities.

Gp \_\_. Full discussion with group of one parent's deep concern that she, her husband and two children have been left in a cramped Council flat, with only one bedroom, for seven years; meeting discussed new ways she might approach the housing authorities to get a move.

Gp \_\_. One participant has a child who is said by nursery teacher (and also assessed as such by E) to be suffering severe coordination and vision problems; mother refuses to accept that child has any problems and would not accept clinic appointment offered by school head; says one must have faith that the child is merely slow in developing; E expressed sympathy with the mother's faith but urges the need for child to attend clinic in order to clarify matters; mother dubious.

6th mathematics meetings

Gp \_\_. A warm sense of community in the discussion between one black and two white mothers; all three parents working on immediate recognition of numbers near 10.

7th reading meetings

Gp \_\_. Mother of simple developmental level reports for the first time some considerable progress by the child; older siblings had started to cooperate in the task and this had apparently 'worked'; mother delighted.

Gp \_\_. Very full and successful discussion of the kinds of problems parents face in sustaining the child's interest in programme activities, particularly in view of competing attractions.

Gp \_\_. Teething baby caused some disturbance.

7th mathematics meetings

Gp \_\_. Mother with 'pushful' relative again came late; useful discussion on the problem after the other participant left for home.

Gp \_\_. Mother (sole attender) depressed and admits to desire to beat obstinate child over his behaviours at home (unrelated to programme activities); matter discussed at length; programme hardly discussed. (E heard later of

appreciative comments by mother to nursery teacher, of opportunity to discuss her worries at several meetings where she had been the sole participant; reception teacher said that child's behaviour problems had eased considerably and the enuresis had stopped.)

Gp \_\_. Only one (Asian) mother present; clearly embarrassed at finding herself in that position; E tried to handle matters tactfully, did not look directly at her.

#### 8th reading meetings

Gp \_\_. Mother reports traumatic effects of visitor who had called, had found her and child engaged on programme activities, and had declared "This girl should not be doing that kind of work now - she should be playing"; statement seemed to have affected child's willingness to do the activities; E suggests leaving matters a while, then trying again later.

#### 8th mathematics meetings

Gp \_\_. Mother reports considerable progress by daughter (who had recently entered infant school); programme has helped her to master base 10 concepts.

Gp \_\_. Letter from absent parent who apologised for absence due to illness; said she or her husband would like to be given the final agenda ideas 'any time, any place'; E visited the home after the meeting.

School \_\_. The mathematics teacher, who had asked to see a summary of the programme and some of the materials, said it was well structured and showed a progressive development.

### 6.36 Parents' views on the programmes

At the final meeting of each group, parents were asked to comment on a number of points. Altogether 22 out of the original 25 groups were still in existence, although in 10 cases this was nominal in the sense that only one parent was in attendance. The last 15 minutes of each final meeting was devoted to the following questions, posed informally by E:

What aspects of the programme 'worked' and what aspects did not 'work'?

Was the programme too fast, or too slow?

How does one adapt such a programme to parents whose children develop either faster or slower than others?

If such a programme were to be run again, for other parents, are there changes you would like to see in the programme?

Do you feel that the (two-weekly / six-weekly) meetings were of the

right frequency, or would you rather have attended more meetings and had less material at each meeting, or fewer meetings with more material?

Are there other comments you would like to make, which could help E to assess the programme?

Specific comments and suggestions made at the meetings can be classified under a number of headings. Because of the varying numbers at the meetings and the difficulty in making quantitative assessments of the strength of feeling on the issues, in most cases no attempt is made to indicate the weight of opinion. Comments made by parents in the final meetings of the two pilot groups are included here.

a. Overlapping with school work?

Some mothers who had started a programme but then given it up, told a continuing participant that they couldn't see the point of doing what the school teacher was going to do in any case. Another view was that many of the activities that the child brought home from school could overlap with the activities in the programme. Was there not a danger that children who had progressed in the programme could become bored at school? Discussion of this point raised the parallel issue of whether ex-nursery class children could be bored in a reception class where many activities were designed for those who had not attended nursery.

b. Value for parents?

One view was that many mothers were already doing these kinds of (programme) activities already. Other views were that parents had been unsure of what to do, but the programmes had given them many ideas; for the first time they had an awareness of how to help their children in these areas. Nearly all the parents said they had learned things and been able to help their children to attain skills which they would never have done by this stage in the ordinary course of events.

c. Interaction between parents

Comparing of experiences with other members was seen as important. The meetings offered regular stimulation and enabled parents to exchange ideas, according to some mothers. The only problems which could arise in this area concerned the shy parent, differences of personality, and the formation of cliques; these might be among the reasons why some mothers had opted out of the programmes - other parents might have done so because they did not want to put themselves out. One person suggested that perhaps E should have been more formal at times, in order to cut out some of the unnecessary chatter in

which participants indulged.

d. Children's individual needs

Although this matter was not raised to any serious extent in the final meetings, it arose often during the course of the programmes. Various parents argued that it was a pity that E's approach could not be more personalised and adapted to their particular child's needs and level of development; but they recognised the difficulty of this proposal in a situation where the children were at different ages, levels and interests. Another difficulty seen by parents was that too much individualisation would focus attention on those children who were only progressing slowly, and possibly cause embarrassment to some mothers.

e. Children's particular problems

Among the points raised was the question of children's changeability; parents had to be sensitive to their children's differing needs, for on some days they asked for the programme games and were very keen, and on other days they were bored or unwilling even to try any games. Another point was that if meetings were held only at long intervals, parents could be left with problems on their hands which they would rather try to resolve quickly.

f. Frequency of programme meetings

The decision to run two-weekly and six-weekly programmes was made in order to see whether there were any benefits attached to less concentrated or more concentrated schedules. A related issue was the notification of meetings; notices were posted to six-weekly participants, whereas two-weekly participants were expected to remember the meeting days, assisted only by the posters displayed on the day of the meetings. Factors such as convenience and forgetfulness were among the reasons why parents might favour one or other kind of schedule (random allocation naturally allowed no prior choice in the matter). Although ostensibly an administrative matter, attendance and the nature of the meetings held at different frequencies might well be of some importance in terms of programme effectiveness.

The following arguments were mentioned by parents favouring one or other of the alternatives below. (The basic choice presented was between frequent meetings, weekly or fortnightly, for about an hour, or less frequent monthly or six-weekly meetings lasting a whole nursery 'morning' or 'afternoon', that is, over two hours per meeting. Associated with this choice were two further considerations: the longer meetings would present the same amount of material as would normally be presented in two or three of the shorter meetings; and tea would be offered at these longer meetings.)

Weekly meetings: a few parents, mainly from the pilot (fortnightly) meetings, favoured the idea of weekly meetings as they helped to keep people 'on their toes' and remind them to keep up with the programmes.

Fortnightly meetings: about half the parents favoured continuing with fortnightly meetings, rather than having fewer but longer meetings. The burden of the comments here was that parents would be inclined to forget or slacken and would not carry out the same amount of activities with the children, if the meetings were only held at longer intervals. There would be a tendency to 'cram' more activities if they were only given monthly, say; also it was difficult to concentrate if one was given too much at any one time. Two weeks was a happy solution, offering continuity and stimulation. Moreover, many parents relied more on being shown things, and remembering them, as occurred with fortnightly meetings, rather than having to rely more on reading the explicit programmes needed by the monthly or six-weekly groups. "I can still picture in my mind how you demonstrated things", said one parent.

Monthly meetings: a significant minority of the two-weekly parents favoured monthly meetings. It was argued that people had many commitments and one morning or afternoon a month was less intrusive than fortnightly meetings. Two hours would give participants more chance to have constructive discussion of the programmes. One advantage, cited by a number of people, was that with the meeting occupying a whole nursery class period, they would not have the burden of walking home after the shorter fortnightly meetings, before returning to fetch their children. Provided parents were reminded by letter or note in advance, there would be no problem about remembering the days. Several mothers felt that it would not be too difficult to space out the work if they were given two fortnightly programme sheets at the same time. A major argument voiced by some parents was that a month would enable them to see more progress in their children and they would have more to report on at the meetings.

Six-weekly meetings: most of the relatively small number of parents who attended the six-weekly programmes favoured this period. They said that the programme sheets enabled them to follow the work that had to be done in the intervening period. Although they could forget points that had been demonstrated at the meetings, they were reminded of these points by re-reading them later in the programme sheets.

#### g. Attendance difficulties

On the problem of forgetting, only relatively few parents claimed that it posed no difficulties. One said: "If a mother is aware of the programme and doing something with the child regularly, it is easy to remember". But more parents pointed to the difficulties: for example, there were often mix-

ups as to whether the next fortnightly meeting was the coming Monday, say, or the Monday following. Others felt that a monthly meeting might lead to yet more problems of forgetting, even if notices were sent out. One practical suggestion was that notices could be left, a few days beforehand, in envelopes for parents on the nursery class board. Many parents wanted to be notified no more than three or four days in advance.

#### h. Level of the programmes

This was seen as a key issue because the setting of activities at an optimum level would be likely to have optimal effects on the children's later attainments. It would also affect the parents' sense of achievement, if the programme was at the right level, compared with a sense of failure if it was too difficult or boredom if too easy. It was inevitable that answers to this question would be biased in two ways. Parents remaining in the programme were those who felt they were getting something out of it, and perhaps some of those who left might have done so because the level was too advanced for their particular children. The other reason for bias was that parents with bright children would clearly find the programme acceptable or too easy, and the reverse would be found with the parents of slower children, so that one might really need to relate comments to some criterion like the children's cognitive levels - an impossibility.

At the two extremes of responses were the parent of a bright child who commented that the level was fine, the child had coped well and she was glad that she had been given the full programme to date, and the parent of a somewhat slower child who said that "Maybe it was a little too much, but if you keep the materials you can do the work when you have the time to do it all".

Various parents argued that it was right to set the programme at some kind of mean level, knowing that it would be too fast for some and too slow for others. One mother wished that it were possible to give a programme with less work for the child who was not doing so well, and more for the one who was progressing fast. Another parent commented that it was an advantage that the programme could be followed at different levels, unlike the situation in some classrooms when all children were expected to work at the 'average level'.

Specific quotations might indicate the tone of the comments:

"I know to what level my child can work and therefore I prefer to get all the material and give it to him when he is ready, rather than getting a smaller amount.... It is important for mothers to pick among the items and decide what they want to try out on their child, and what they will delay until later."

"As long as one understands what is being said, and how it is organised - in what order - then I am quite happy with the level of the programme

achieved."

"I can only go as fast as the child can take, and (in terms of) how much time I have as a mother. The (regular programme) sheet seems rather daunting, but when one looks at it and sorts it out, it's not so bad."

"It seems a lot, but it's not too bad provided one doesn't have to keep up with the programme. If it had been within a set time, it would have been bad."

"A lot of things have not yet been touched, but I can go through it (the programme). At least I know where it is."

"I would prefer the programme (level) as it is; I can use this work when he goes up to school. (It's really been) a parent education programme."

"It was an ideal amount of work given with my child - who is just an average child."

#### j. Structuring of the programmes

The relatively formal structuring of the programmes was a deliberate feature aimed at improving the effectiveness of the parents' activities. One parent, herself a mathematics teacher, commented that she found the mathematics programme very logical. Another parent found the ideas good, more so than the materials which could not be strictly adhered to by everyone. The quantity of materials provided was favourably commented upon by many parents, particularly because they considered it afforded more choice of games. (This could in fact be seen as indicating a partial avoidance of the structuring element of the programmes.) Parents considered the demonstration of activities to be very necessary; the materials themselves could not be understood without the memory of the demonstrations. Some parents did find problems in linking the different activity sheets with the individual agendas for particular meetings; it was suggested that the sheets should indicate the agenda numbers, and not merely the other way round as at present. It was generally agreed that the accelerated programme schedules for the final three meetings had been more difficult to follow. (This point had been made to parents earlier by E, when explaining the reduction in meetings from fourteen to eight.)

### 6.37 Some methodological considerations

Later sections deal with the 'parent programme' intervention variables and their characteristics. At this stage it is only necessary to refer to a few points of methodological importance.

When following any criteria for allocation into programme and non-pro-



programme groups there are marginal cases where ad hoc decisions have to be made. Parents who accepted the invitation to attend programme meetings and who started are clearly identifiable. Parents who were in full-time employment and who offered this as a reason for their inability to attend are equally identifiable. The three parents who expressed hostility to the ideas underlying the programmes were placed in the refusal group, whether or not they were in employment or available to attend meetings.

It was the small number of remaining parents who posed the problems of allocation. Some parents who worked part-time said they would try to attend; a few did, and a few others reported later that their employers would not give them time off (such parents were usually shop assistants). All those parents in part-time employment who said they would try to attend but did not turn up to any meetings were added to the working parents group.

A few other parents were also added to this group. One was an immigrant mother, from a Continental ethnic group; the home environment was warm and supportive but the mother could speak no English; there was an older daughter who was already reading English stories to the target child. This mother was classified as part of the working parents group - it appeared more appropriate than the refusers group. Another mother, from a different ethnic group, said she was interested in the programme and would attend meetings when it suited her, though not regularly; E emphasised the importance of regular attendance but the mother would not be moved. Finally E said that he did not feel it was right to keep a place open in a group since the sequence of activities was an essential part of the programme. This parent was also classified as part of the working parents group. Despite E's seeming 'rejection' of the parent, relations remained good when they met at the nursery later.

Reasons for putting other cases into the working group, such as the child minder with four or five children in her care who attempted to attend but never got as far as the meeting venue, have already been described. Another mother who accepted the invitation but did not attend was known to live far from the school; the possibility of having to travel to and from the nursery twice in the same day was clearly too daunting, and she was accordingly placed in the working group rather than in the non-attenders group. Similar principles were applied to a few other parents, such as mothers near to giving birth.

It was only those parents for whom there was no clear and justifiable reason for non-attendance, in the view of E or the nursery teachers, who were allocated to the group of accepters who did not attend any meetings, this group being labelled 'non-attenders', in contrast to the 'working parents group'.

One criteria for the effectiveness of the programme would have been an

assessment of the parents' post-programme home behaviours, similar to the assessment carried out during the original interview with parents. There would have been little point in asking programme parents to pass judgement on their own behaviours. On the other hand, for E to re-interview parents, or even for an outsider to have been asked to undertake this work, would almost certainly have brought answers from programme parents which reflected in some degree their desire to show that the programme had influenced their behaviours. This would be a natural response. Only an in-depth interview, of considerable length, with a skilled questioner seen to be unrelated to the programmes given by E, would be likely to obtain an unbiased view of current parent behaviours.

Accordingly the only real judgement could be in terms of assessing child performance on the post-tests, taking into account pre-test child performance and cognitive levels as well as initial parent behaviours.

The basis of the intervention scoring has already been described - 2 points for attendance at the first meeting of one of the two programmes and 1 point for subsequent attendances, with 2 further points for having received the main set of programme cards. Parents who missed a meeting but were subsequently given the details and materials during a 10 or 15 minute discussion with E were credited with a  $\frac{1}{2}$  point for that meeting; in some cases, when it was possible to give a fuller supplementary meeting (in the home, for example), this was scored as 1 point.

The intervention variable score was determined by multiplying each parent's total points by 4. Thus reading or mathematics attenders would be given a 'reading meetings' or alternatively a 'mathematics meetings' variable score related to the level of attendance. Clearly the only sample parents to score on a particular meetings variable would be those who had attended group meetings for that programme. The statistical handling of these restricted variables within the total sample model will be discussed in later sections.

Since a number of the meetings had to contend with varying levels of noise and physical disturbance - inevitable in the context of the nursery foyer where many of the meetings were held, or in the context of noisy crying or other disturbance from toddlers within the parent group itself, the diary entries for meetings (where noise and disturbance levels had been noted) were examined and a 'weighted meetings score' determined. This was on the basis of separate variables in which the reading (or mathematics) meeting attendances were scored not as an equal 4 points for every meeting point, as above, but in relation to the noise and disturbance levels at each meeting. Thus the lowest score, for a meeting beset by maximum disturbance, was fixed at 2 points; the highest possible score for a meeting with little or no disturbance was 5 points.

These weighted scores were combined to form the 'weighted reading meetings' or 'weighted mathematics meetings' variables. It should be emphasised that this weighting was not in terms of some unquantifiable measure of the 'success' of an individual programme meeting, but simply in relation to the degree of environmental interference, as judged by E, with the advantage that a single judge was applying the criteria across all the meetings, subjective though the criteria inevitably were.

One last methodological question concerns the possibility that the programmes were, perhaps unawares, geared specifically to the post-tests. In the reading area there are a few words which were inevitably incorporated into both the programme and the Infant Reading Test, such as Mum, School, house and brother. These words are in widespread use in most reading schemes. E was aware of the danger of 'teaching to the test' however, and was particularly careful not to cite any specific I.R.T. environment words when encouraging parents to extract such words from their own environments and use these for teaching simple words to their children. Bus stop (again widely found in reading schemes) was perhaps the only environmental word which was specifically cited at programme meetings and which also appears in the I.R.T. One other such word, Woolworths, appears in one of the readers and also in the I.R.T. The only children who scored an 'isolated' point in the recognition of Woolworths were the child of a non-attender and the child of a mathematics group parent.

In the mathematics area there was one agenda item dealing with conservation of quantity, but for this parent-child activity (which only a few parents reported having tried out) it was suggested that an equal number of Unifix blocks be placed in two jugs of different diameters. This was deliberately made as different as possible from the Piagetian conservation of number test administered by E. Other items on the various mathematics agendas suggest only a remote or indirect link between the programme content and the concepts and numeracy items appearing in the mathematics or Piagetian tests.

#### 6.38 Lessons for future programmes?

The shortcomings that became apparent during the running of the programmes led to the consideration of possible ways to avoid these problems in future intervention projects of this nature. A number of the ideas came from the parents themselves. There are of course many other possibilities for improvement beyond what is outlined below.

1. The overriding.../

1. The overriding importance of giving parents written or other advance notice of meetings, two to four days prior to the meeting date, showed itself with the many reports from parents who forgot the dates of the next meeting and had already planned other activities by the time they reached the nursery door with their child.
2. A second issue of major importance in this type of project is the need for some partial individualisation of the programmes. If the ideas of many parents in this project were to be accepted that there should be monthly rather than fortnightly meetings, but covering an entire nursery session, it would be possible to divide the period into three parts: one-third for a discussion of behaviour problems and other developmental issues seen as important by parents; one-third for the presentation and discussion of programme ideas and giving out programme materials for the coming month's activities; and the last third to be organised so that while the group discusses and clarifies the application of what has been presented, individual parents will in turn discuss the child's progress or problems with the programme organiser, free from the constraints of discussing such matters within a group situation.
3. Reliable procedures for bringing interested parents up to date with a programme, if or when they unavoidably miss a meeting, are essential in terms of a structured sequence of activities within the programme. This was done on an ad hoc basis in the present project but it would be preferable to have a specific procedure known to parents and the organiser and accepted in advance, whereby there would be some extra individual time offered to a parent who had missed a meeting.
4. The child's support for the goals of the programme might be enlisted more strongly if each child was given an individual decorative card listing some of the major attainable goals in the programme; this card could be displayed on a wall at home and as the child achieved a particular target performance the card would be marked by the parent accordingly.
5. Reading group parents who had not started using their local library by the time the second programme meeting was held would be inveigled into accompanying the organiser, during or just after this meeting, to the nearest local library to make an initial exploration or 'guided tour' of what the library has to offer. This suggestion is related to the minimal success achieved in the present programme in persuading parents to take the bold step of joining a library or at least going there for a visit.
6. In the reading programmes there might be more activity papers specifically geared to learning the words for body parts, important home items and key environment words. While it would be desirable for parents to accumulate these words themselves, parents with relatively low levels of literacy

or verbal sophistication may find it difficult to think up such sets of words themselves.

7. The mathematics programme might likewise stress certain major goals, such as the recognition of numbers from 1 to 10 and the ability to count and match these and the understanding of certain basic terms and concepts. More activity papers could be based on such skills and more repetition work provided, in new forms, to ensure effective learning in some key areas of the programme.

8. The possibility might be considered that this kind of programme, running perhaps over seven monthly meetings within any one calendar year, could be provided for all parents of nursery class children in the more disadvantaged areas. This might entail several mornings and afternoons of meetings each month at the schools with larger nursery classes, and fewer meetings for the smaller schools. The meetings, preferably in small groups of six or seven, would be sufficient in number to offer parents a choice of days for them to attend. Given a conviction that such programmes are of use in the early educational development of children, infant teachers from the schools themselves or alternatively the nursery teachers might run these parent meetings. It would help to establish closer liaison between what the parents were doing with their children and what the teachers hoped to be able to do when the children moved into reception class.

9. The recruitment of 'parent group leaders' at each school might also be considered. The experience in this project was that there are a few parents in every nursery class population whose qualities of rapport with other parents and insight into the parent role in early education are such that much of the organisation and motivational work for these programmes could be entrusted to them, leaving the teachers with the main responsibility for providing quality programmes.

10. Within a non-research situation, where the comparative performance of reading versus mathematics groups is not sought, it would be reasonable to offer parents a combination of an early reading and early mathematics programme over the seven meetings, in view of the educational importance of both areas. For the parent it could be an avenue to practical awareness of these issues, and for the child a means of developing an early educational interest in which school, parent and child are interlinked.

#### 6.40      Data, sub-sample and group characteristics

This section examines the data at their most basic level and looks at the characteristics of sub-samples and programme groups, to give an overall picture of the statistical material which is to be employed in the analyses presented later in this chapter.

The first of the following sub-sections deals with the broad characteristics of the sample data for over 50 continuous variables which have been derived for each child and parent combination. The distributions of the variables are examined for departures from the normal and reference is made to the more detailed distributional statistics listed in Appendix D 3.

The next sub-section examines in some detail a selection of sub-samples which can be identified within the main sample; these are related to population characteristics rather than to the research characteristics imposed on the sample. Mean scores are compared on a number of key variables and the probabilities of the differences between sub-samples are measured to give an indication of how similar or disparate these groups are when assessed on such variables

The third sub-section is of particular importance because the sample is divided into the various programme and non-programme groups and the performance of each group examined in terms of some 13 key variables. The probabilities of the differences between the groups are determined.

The fourth sub-section brings together some of the findings from the previous sub-sections and discusses the experimental design and its limitations in terms of those findings.

#### 6.41 Broad characteristics of the data

Detailed figures on the data assembled in this study and the statistics derived from them appear in the appendices D. Appendix D 1 provides a summary of these appendices. Appendix D 2 lists the raw scores from the four sets of test, interview and programme data. Appendix D 3 lists the means, standard deviations, skewness and kurtosis of each of the continuous variables, while a small selection of histograms from the more interesting or unusual variables appears in Appendix D 4. To facilitate comparison, the histograms are based on standard scores.

There were 52 continuous variables and 8 categorical variables regarded as having sufficient relevance and discrimination to be incorporated in the analyses. To clarify the picture, all these variables are grouped in table 19 overleaf, together with the variable coding or acronyms used in the path analyses and data appendices.

In the following pages the performance of each set of variables will be reviewed very briefly, with more attention paid to variables whose distributional characteristics are unsatisfactory or worthy of comment. Skewness and kurtosis statistics are of value in this interpretation. Essentially, skewness measures the degree to which a distribution approaches that of a fully normal curve, while kurtosis provides a measure of the degree of peakedness or flatness of the distribution - whether it is too narrow or too wide relative to the normal. Although histograms were obtained for all the variables, using a specially modified computer programme to give a clearer picture, only a selection are reproduced in Appendix D 4 and elsewhere in the report. However, all the histograms were studied and considerable use was made of the insights they provided about the variable frequency distributions, especially when carrying out the more complex analyses.

Prior to using the data from each set, missing scores were replaced by the means of the variable in question. There were 6 missing values out of some 2,400 scores in the nursery battery, 4 out of some 2,000 scores in the reception battery, approximately 45 out of 5,600 item scores in the interviews, and none out of the 800 scores in the post-tests. It was possible to keep the missing value totals at a low level in the tests since frequent checks on score sheets enabled any overlooked tests to be administered a day or two later. Replacement of missing values in the interview coding was more difficult, because usually only one item would be unscored in a multiple-item variable. Where possible, the missing value was based on the level of the remaining items in the particular cluster, or if not possible a less favourable score was entered for high scoring parents and a more favourable score for low scoring parents.

Table 19. Summary of main variables used in the analyses

<u>Nursery attainment variables</u>		
Reading related:	English Picture Vocabulary Test	EPVTS
	Reading awareness	Rdgaw
	Infant Reading Test	IRTSN
Mathematics related:	Mathematics numeracy assessment	MatnN
	Mathematics concepts	MatcN
General:	Piagetian tests (composite)	PiagN
<u>Nursery cognitive variables</u>		
Cognitive skills:	Wechsler Pre-school and Primary Scale of Intelligence:	
	Information sub-test	InfmN
	Sentences sub-test	SentN
	Picture completion sub-test	PicoN
	Block design sub-test	BlocN
Meta-cogntv. performance:	Rhythmic tapping	RhytN
	Matching familiar figures	MFFIN
	Bender Gestalt	BendN
	Self-picture test	SelfN
	Distractibility (negatively scored)	DistN
Other contributors:	Sex	SexN
	Age at nursery tests (half-months)	Ageny
<u>Reception (mid-test) cognitive variables</u>		
Cognitive skills:	Wechsler P.P. Scale of Intelligence:	
	Information sub-test	InfmR
	Sentences sub-test	SentR
	Picture completion sub-test	PicoR
	Block design sub-test	BlocR
Meta-cogntv. performance:	Rhythmic tapping	RhytR
	Matching familiar figures	MFFIR
	Bender Gestalt	BendR
	Self-picture test	SelfR
	Distractibility (negatively scored)	DistR
Other contributors:	Sex	SexR
	Age at reception tests (half-months)	Agerc



Table 19 (continued)

<u>Post-test attainment variables</u>		
Reading related:	Infant Reading Test	IRTSP
	Southgate Reading Test	South
	Daniels and Diack (sentences) Test	Diack
Mathematics related:	Mathematics numeracy assessment	MatnP
	Mathematical concepts	MatcP
General:	Piagetian tests (normal failure score)	PiagP
	Piagetian tests (successive fail.score)	PiagX
<u>Post-test cognitively related variable</u>		
General contributor:	Age at post-tests	Ageps
<u>Parent-home environment variables</u>		
Reading related:	Reading behaviours	Rdgbh
	Language environment	Lange
	Parent reading attitude	Prdga
Mathematics related:	Mathematics behaviours	Matbh
Other contributors:	Parent-child cooperation	Pcoop
	TV viewing time (negatively scored)	TVtim
	TV control behaviours	TVbeh
<u>Parent programme variables</u>		
Reading related:	Reading meeting attendance	Rdgmng
	Reading meetings: weighted score	Rdgtwt
Mathematics related:	Mathematics meeting attendance	Matmg
	Mathematics meetings: weighted score	Matwt
<u>Other variables</u>		
Needs (Maslow criteria):	Need for security (nursery assessment)	NdscN
	Need for esteem (nursery assesst.)	NdesN
	Need for security (reception assesst)	NdscR
	Need for esteem (reception assesst)	NdesR
School experience (age-adjusted, using standardised scores)	Time in nursery (after nursery tests)	Tinur
	Time in reception (up to post-tests)	Tirec

Categorical...../

Table 19 (continued)

Categorical variables

Sex (included in above groupings)  
School  
Morning/afternoon nursery session  
Ethnic group  
Disadvantaged/advantaged socio-economic status  
Views on purpose of nursery education (coded into 6 categories)  
Programme and non-programme group allocations  
Moved but traced to new schools for final tests

6.411 Pre-test data (nursery)

The distributional characteristics of the six variables in the nursery attainment set and the 11 variables in the nursery cognitive set are given in Appendix D 3. One of the variables, Sex, is excluded from further assessment in this sub-section, beyond noting that for the sample as a whole there is a modest preponderance of girls.

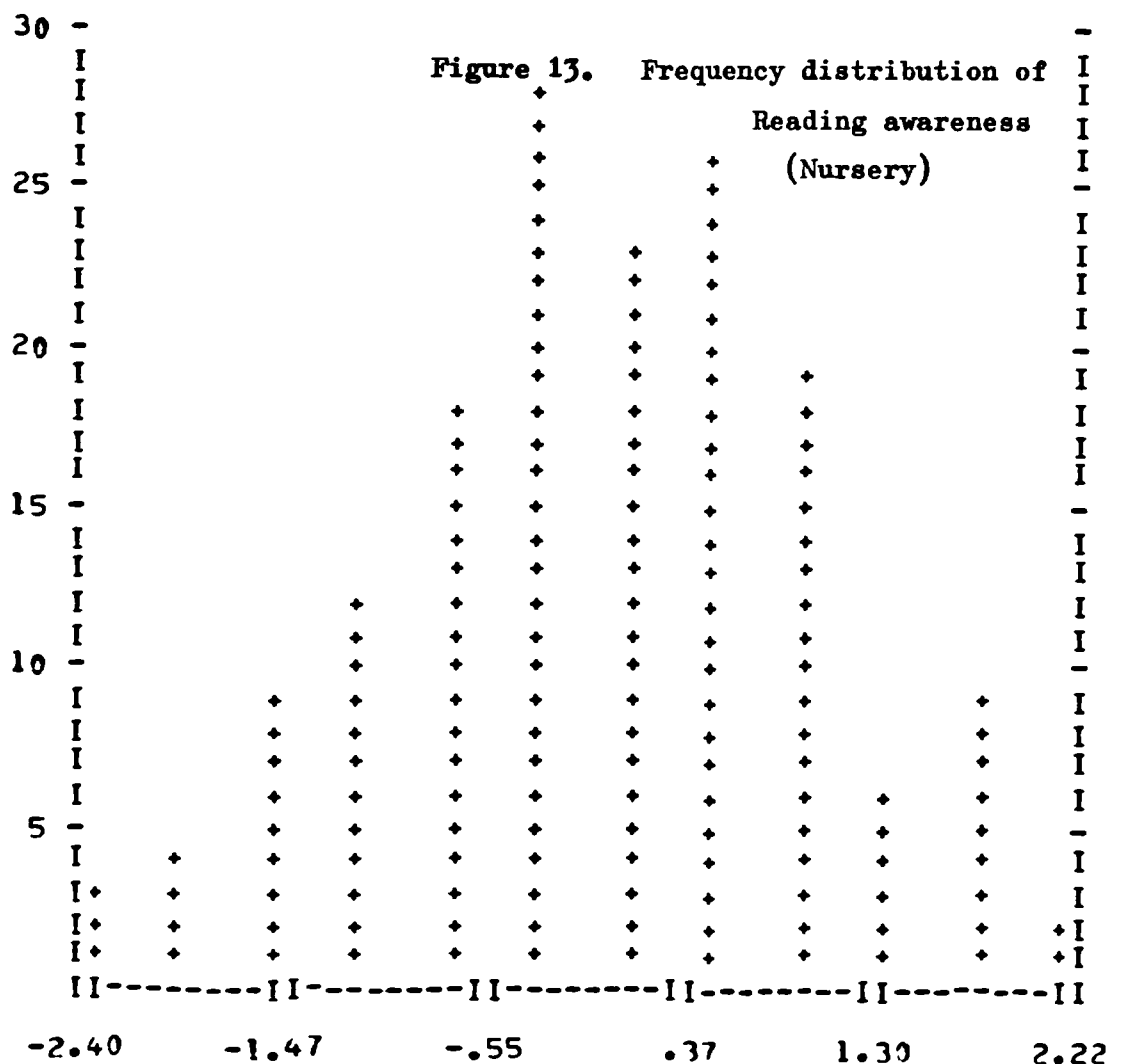
Of the remaining total of 16 variables there are four whose characteristics differ considerably from the normal distribution. Although attention is concentrated here on these 'awkward' variables, it should be emphasised that the regression methods which form the main approach to analysis in this study are not unduly affected by non-normality in the predictor variables. While an unsatisfactory distribution can nevertheless be a reason for the poor performance of a variable in the model, it is more likely that a fundamental conceptual relationship or lack of relationship with other variables will play a far larger part in any performance within a model than does a moderately unsatisfactory distribution.

The histogram of the Infant Reading Test is reproduced in Appendix D 4. It shows that at the nursery level this test is considerably skewed, as borne out by the skewness statistic; but it also shows that there is a reasonable spread of early reading scores if the very youthful age of the children is taken into account. The author is unaware of any other reading test which can measure incipient reading levels at the age of four. It can be noted that some 65 per cent of the children score within a peaked 'normal' distribution of between 1 and 6 points, while the remaining 35 per cent are strung out in the direction of higher scores, suggesting that home activities for this minor-

ity tend to emphasise the acquisition of early reading skills.

Rhythmic tapping is also considerably skewed, with many of the White children in the sample unable to copy the simplest tap patterns; in contrast most Black children achieved well on this item. Another highly skewed distribution is distractibility; this, however, is an artefact of the testing protocol; the majority of children were found to show little or no sign of distractibility during the testing and thus scored at or near ceiling level (this variable being negatively scored).

The fourth problem variable in this set is the composite Piagetian test, whose profile at the nursery level is very highly skewed, offering in fact a J-distribution. The histogram (Appendix D 4) shows that a considerable number of children (about 40 per cent) failed to achieve any score. This would be in line with the hypothesis that at the age of four there are many youngsters who have not yet grasped the concept of same/different. Further comment on this test appears under the discussion of post-test data.



The remaining variables are fairly satisfactory. Examples appear in D 4. Some variables, such as Reading awareness (whose histogram appears above), Mathematics numeracy, Bender Gestalt, Matching familiar figures and Self-picture

(draw-a-person), have sound profiles, indicating that the minor modifications which have been introduced into the scoring of several of these tests (as described elsewhere) have not affected their distributional properties. One variable, E.P.V.T., has a fairly strong negative kurtosis (flatness), but this is partly due to the wide age range of administration of the nursery tests. The behaviour of the WPPSI sub-tests will be discussed in the next sub-section.

Overall it may be said that on the attainment side the nursery tests present a picture of the considerable differences existing in children's 'academic' levels even at the age of four, before they have experienced any formal schooling. The positive skew in the verbal measures (attainment and cognitive) is more evident than in most of the non-verbal measures. This is in accord with the hypothesis that verbal skills are likely to reflect the parent-home environment more than do other kinds of performance at this age.

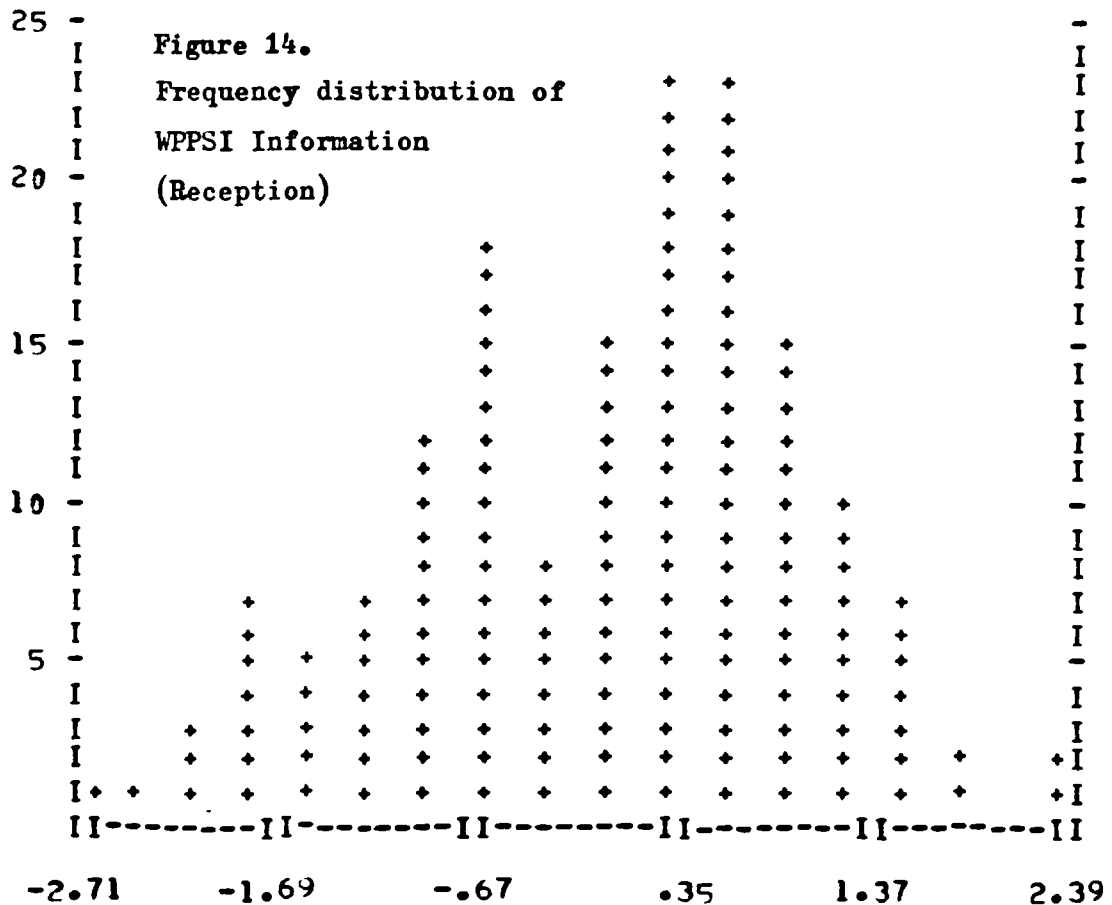
#### 6.412 Mid-test data (reception)

The profiles of the 10 continuous variables in this set show, on the whole, more satisfactory distributions than when the same variables were administered at the nursery level. This is to be expected, since there were fewer children scoring at near zero levels on any of the tests.

The highly normal histograms of some variables, such as Matching familiar figures and Self-picture, are reproduced in D 4. As with the previous battery, distractibility remains highly skewed in view of the scoring method; the skew in the rhythmic tapping profile is somewhat reduced, reflecting the larger number of White children who can now master the simple rhythmic skills.

The distributions of two of the WPPSI sub-tests are not altogether satisfactory. The Sentences sub-test continues to show a fairly high positive skew, with the added problem that there now appears to be a second peak developing among the higher scorers. As these scores are not age-corrected in terms of the Wechsler protocol it could be argued that their skew merely reflects the age distribution of the sample. However, a study of the nursery and reception age histograms (reproduced in section 6.60) offers no real support for that hypothesis. It may well be that there is a kind of barrier within the Sentences items or in their administration. After this barrier has been overcome the test may be measuring some other skill than that measured in the early part.

The problem presented by Block design is more worrying. For comparison the profile of the Information sub-test (reception level) is reproduced on the next page. This is clearly a variable with an eminently normal distribution. In contrast the Block design (nursery) and Block design (reception) histograms are reproduced on the page following, partly superimposed on each other. In

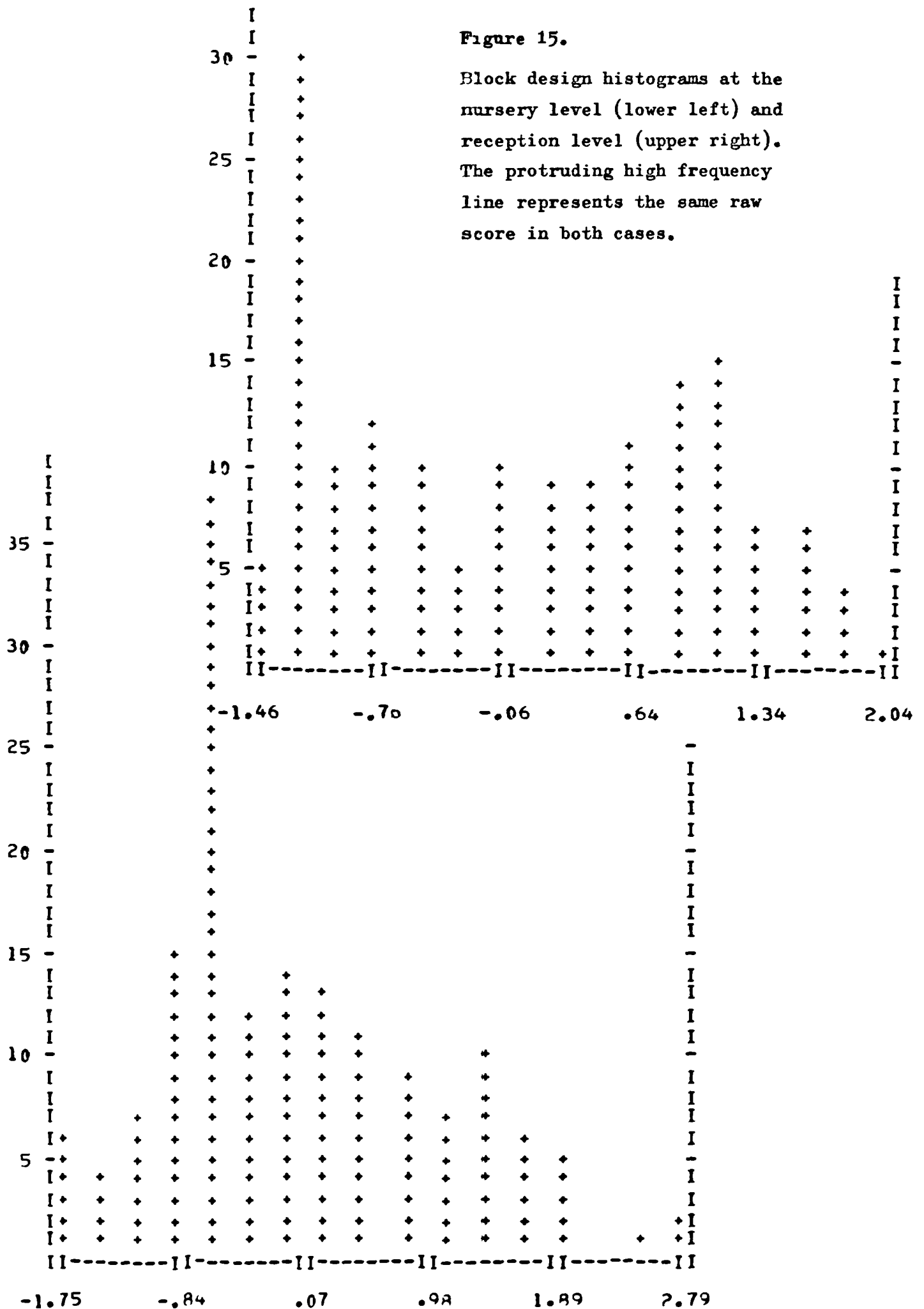


both cases the protruding frequency of 30 or more cases occurs at the score point of 4. (For BlocR, the reception variable, there were only a few children still below this critical point.) Although Sattler (1974) points to the difficulties caused by the administration protocol for the Sentences sub-test, this flaw in Block design appears potentially more serious. The nature of the problem makes it unlikely that it is due to experimenter error; rather there appears to be a considerable jump between the skills needed to cope with the first and second items (arranging three square red and white blocks in a patterned sequence) and the skills required to score on the third item, namely to arrange a simple diagonal pattern. (The issue of the children's ability to handle diagonal patterns is discussed very briefly in Appendix D5, under Bender Gestalt.)

Despite these scoring problems, and the evident loss of discrimination power for the relatively large number of cases who are barred at score 4 - the full mark for the first two items - there are good grounds for retaining the data from the Block design test because of its general performance as a measure of non-verbal skills.

Figure 15.

Block design histograms at the nursery level (lower left) and reception level (upper right). The protruding high frequency line represents the same raw score in both cases.



6.413 Post-test data

Appendix D 4 reproduces six out of the seven histograms in this set of data.

Among the three reading variables, Infant Reading Test has largely overcome the skew problems that made it less satisfactory at the nursery level. The small remaining skew is no longer serious within the distribution as a whole. Although skewness and kurtosis statistics seem to suggest that Southgate's profile is not too wayward, in fact there are 25 children scoring zero on this test; the frequency pattern for the remainder is satisfactory. The Daniels and Diack results present the worst problem; 74 children, nearly half the sample, score zero on the test, despite the modified method of marking which provides a much better distribution for the remainder. The standard deviation is half as large again as the mean for this test.

There are good reasons for not excluding the variable from the analysis. The test was shown to have useful characteristics in the author's previous research (Barker, 1976). It measures the ability to read sentences and, implicitly, to contribute to reading performance 'from behind the eyeballs' (Frank Smith, 1973). It is thus a higher level reading skill than that which is tapped by I.R.T. or Southgate. Since the variable will not be used on its own as a criterion, its grave lack of normality is not a bar to its use as a criterion variable alongside the other two tests. Although the three tests, in conjunction, will still present a somewhat sharply skewed distribution for this composite reading attainment measure, the combination has enough normal features, apart from its wide range, to serve as the main criterion of post-test performance. (The weighting of these three variables and alternative methods of combining them in a path analysis model are discussed later in this chapter.)

The two mathematics measures have adequate distributions. Numeracy is virtually normal while the Mathematical concepts variable shows a small negative skew, with 8 children scoring at ceiling. Clearly this variable could have benefited by a few more items at the top end of the scale. The unexpected importance of Mathematical concepts as a criterion variable is discussed later.

The two Piagetian measures merit particular attention. The histograms for both the post-test Piagetian variables appear in Appendix D 4, superimposed on each other. PiagP has a distribution very similar to the pattern for the nursery measure, PiagN, and also to a mid-test measure on the same test (the mid-test variable has not been incorporated in the analyses, since it would have added unnecessary complexity to the path analyses, with no particular conceptual

justification for its inclusion at the mid-point).

The basic problem with the original Piagetian measure lies in the conceptual stance on which the protocol was drawn up for its scoring. The theoretical position taken, in line with Piagetian thinking, was to regard the ordered sequence of tests - same/different, conservation of number, seriation, and multiple classification - as patterned on the normal conceptual development of Western children at the ages of around four to six. Accordingly failure at a certain point was interpreted as conceptual failure marking a point just at or just beyond the child's level of Piagetian development. However, experience with administering the test to some 250 children (sample and pilot) at the nursery stage, and to 200 children at the mid-test stage, suggested that many children might in fact be able to perform at higher levels than their initial failure point. This was not seen simply as a difficulty in understanding a particular word formula but as a real inability to grasp the principle underlying conservation of number, for example, and yet having an ability to perform seriation tasks and even on occasion multiple classification. (Theoretical and research evidence on this Piagetian model have been discussed briefly in section 4.31.)

Accordingly it was decided to administer the Piagetian composite test in the normal manner in the post-test battery, marking the immediate failure point, but then continuing with the test until there had been failure at two successive levels, and deriving a new score, PiagX, for this wider protocol. A comparison of the frequency charts provides the justification for this alternative measure. Although there is now a fairly serious negative kurtosis, the skewness has been drastically reduced in the case of PiagX and there is a far wider distribution of scores.

The importance of this alternative measure was not foreseen at the time the protocol was adapted. It was simply added to the test battery as a matter of research interest. However, the early path analyses of the mathematics programme sample brought out the value of the second measure. Speculation on what these Piagetian tests may represent will be discussed more thoroughly in the appropriate sections.

#### 6.414 Parent-home interviews and programme scores

In view of the difficulty of quantifying the parent-home environment on the basis of an interview schedule, and the fact that most of the variables are combinations of items which are either dichotomous or range over a few points, the distribution patterns of most of the seven variables in this set are reasonable. Four of the histograms appear in D 4.



The double peaks in Language environment and TV time are clearly unsatisfactory, but these could well be artefacts of the item combination process. Television (control) behaviour covers only 4 points, with nearly all the scores on the middle two; this may be one key factor in the poor performance of this variable in the analyses. Parent reading attitude has a heavy negative skew but also a wide spread. Although at the time of the interview E's impressions suggested that he was being told what some parents thought was the right attitude to express (on the three items used to quantify the variable), rather than what they really felt about the issue, the correlation of .51 with parent reading behaviours does suggest some validity for this item.

The programme meeting variables each have a distribution that almost defies statistical definition. Rather than there being a random pattern of loss due to a myriad of individual causes, there appeared to be sizable 'jumps' caused by initial loss of interest, periods of unusually bad weather, or children moving up to reception class followed by (some of) their parents ceasing to attend meetings in order to take up full-time employment. The only usable criterion here is the relative size of the means and standard deviations; for all four variables this is satisfactory, the ratio being well over two in each case.

#### 6.415 Other measures

There are three other sets of interval measures which were quantified. Particular features of the distributions of the age scores - the ages at which each of the three batteries of tests were administered - are discussed in section 6.60, on age and time factors. Likewise the characteristics of the two newly-created variables, school-adjusted 'time in nursery' and 'time in reception', are reviewed in the same section.

The remaining set of variables comprises the two pairs of 'needs' variables assessed by nursery and reception teachers. All four variables have what amount to J-distributions, with a sharp negative skew. The scoring method meant that the child with the least need for security or esteem had the highest scores. The sharp skews reflect the fact that only a limited number of children were thought by their teachers to face serious problems of lack of security or lack of esteem. The histogram for one of these variables appears in D 4.

#### 6.416 Features of particular variables

The administration of some of the test variables brought out particular features and difficulties which deserve brief mention. Appendix D 5 presents a discussion of problems or issues raised by the administration of the following

tests: rhythmic tapping, M.F.F., Bender Gestalt, self-picture test, Piagetian test, the WPPSI sub-tests, the WPPSI picture completion sub-test, distractibility, need for esteem, and the age factor in assessing the youngest nursery children.

6.42 Sub-sample..../

6.42     Sub-sample characteristics

The organisation of the research and the limitation of a single E meant that the children had to be tested over a fairly long time span. As E moved systematically from one school to another, with testing starting at the pilot school in late October 1976, continuing at the first sample school in November, and being completed at the last sample school in early March, it was clear that there would be a wide spread of ages over the sample. A small number of those tested in November and December were moved up to reception class in January 1977, whereas a few of those tested in March 1977 were not due to move up to reception until Easter 1978. This gave a total 19-month span of ages in the sample.

Although the wide spread of ages added an interesting source of variation in the sample, in other ways it was an unnecessary complication, particularly in terms of the age-group aimed at with the parent programmes. It is a difficulty that might have been reduced by sampling more schools and using only the children within a narrow age range, instead of testing all those of four years and rising four within each of the nursery classes, as was done in the present case. In the circumstances, however, and with a single E, the total could not be reduced without endangering the basic goal of ensuring a large enough sample.

There were four criteria for examining sub-samples within the population. Each throws some useful light on the composition of the whole sample. The criteria concern the following differences: between the morning and afternoon sub-samples; between males and females; between the one school with an advantaged intake and the five schools with intakes from disadvantaged areas; and between the children of West Indian and African parents, and the remaining sample. In the two latter cases parent behaviours are also compared.

6.421    Morning and afternoon sub-samples

In most of the schools there was a clear impression among staff that the morning was the favoured time. the morning groups were more likely to be filled, and filled early, than were the afternoon groups. Teachers felt that it was the more aware parent, who planned more in advance, who would be likely to apply first and be allocated the session of choice, whereas parents who arrived at the last minute before the start of term were likely to have to accept an afternoon place.

Analysis of 13 of the 16 variables assessed in the nursery battery (as described in section 4.31) bears out the belief that on balance the children attending morning sessions were somewhat brighter. In nearly half the areas tested the mean scores of the 101 morning children were well above the means of

the 75 afternoon children (attrition losses by November 1977 were approximately the same for both groups). Scores that were 5 to 10 per cent higher included matching familiar figures, Bender Gestalt, maths, Infant Reading Test (22 per cent higher), and two WPPSI sub-tests, Information and Sentences. Three scores were virtually the same - distractibility, need for security and need for esteem. In most cases the spread or variance of scores of the afternoon children was higher than that of the morning children, suggesting perhaps a bigger range of ability in the afternoon. (See Appendix D 8 for details.)

One interesting finding was that on two variables, WPPSI picture completion and WPPSI block design, the afternoon children scored higher than the morning mean. Although the differences were only four and one per cent, respectively, they point to a phenomenon which was also found in other areas, namely that whereas the WPPSI verbal sub-tests distinguish between children's levels of relative disadvantage, this is not the case with the WPPSI non-verbal sub-tests used here.

#### 6.422 Male and female sub-samples

Most of the differences between the 84 boys and 92 girls were in the expected direction, namely with the girls showing a higher developmental level. M.F.F., Bender Gestalt, self-picture test, English Picture Vocabulary Test and maths all showed the girls' mean scores to be between 5 and 10 per cent higher than those of the boys. It is interesting that scores on the Infant Reading Test showed almost no difference, against expectations. Of the four WPPSI sub-tests, Information and Sentences showed small differences in the expected directions. However, boys were 7 per cent ahead of girls on picture completion and approximately equal on block design. (Details: Appendix D 8)

#### 6.423 Sub-samples from advantaged and disadvantaged areas

There is no intention to compare the effectiveness of the different schools in the sample, firstly because the necessary data for that were not collected and secondly because the goal of the research was to compare the intervention programmes and not the schools.

What is of material importance, however, is the kind of problem which schools face when their intake comes from disadvantaged areas, compared to schools which enjoy an intake from advantaged areas. The deliberate inclusion of one school in an advantaged area proved an invaluable source of information on this kind of problem.

It is clear that the nursery characteristics of the children, in terms

of cognitive and academic attainment, are in no way created by the nursery classes or the particular schools, but are rather the characteristics developed from the children's parental and home backgrounds. It should also be stressed that the advantaged area referred to was in no way an elite suburb but rather an area of relatively uniform middle-class homes with a very small minority of disadvantaged homes.

While a few variables (Bender Gestalt, self-picture test, distractibility and WPPSI block design) showed little or no difference, most of the other variables showed sizable differences between the sub-samples. The 31 children from the advantaged area were 18 per cent higher on M.F.F. scores, suggesting their greater degree of reflectivity; they showed a 27 per cent advantage in E.P.V.T. scores, a measure of verbal comprehension; a 14 per cent lead in the pre-mathematics score, which includes simple numerical skills; a 40 per cent advantage in Infant Reading Test levels; they were 22 per cent ahead on the WPPSI Information sub-test, which generally reflects the depth of prior verbal experience and learning of information; and had a 37 per cent higher score on need for esteem - in other words, they were less lacking in need for esteem than were the 145 children from disadvantaged areas. (Details: Appendix D 8) It is interesting to note that despite the significantly higher performance on the information sub-test, children from the advantaged area scored only marginally above the rest of the sample on two further cognitive performances (sentences and picture completion) and showed no difference at all on the Block Design sub-test. This contrast may well reflect the richer experiential environment of the homes in the advantaged area rather than any strong 'inherent' intellectual advantage.

In terms of educational policy and governmental resource allocation, data such as these indicate the extent of the differences that exist (on this set of tests) prior to the start of formal schooling; it suggests that much of what is claimed to be inadequate performance by inner urban schools cannot be laid at the doors of the schools, when the intake itself is so totally different between advantaged and disadvantaged areas. It should be remembered moreover that the sample was seen as reasonably representative of the area's children as a whole, and did not represent an isolated pocket of deprivation. (See also section 6.43)

The second page of Appendix D 8 compares the parent behaviours for the two sub-samples. Here parents in the advantaged area showed mean behaviour levels above those of parents in the disadvantaged areas. Although these differences were consistent (with one exception), the only significant difference among the interview variables occurred on mathematics behaviours (26 per cent higher). Television viewing score (negatively scaled) was 24 per cent higher. Reading behaviour was only 4 per cent higher. The one surprising exception to

the trend was language environment, where the mean score for parents from disadvantaged areas was 5 per cent above that of the parents from the advantaged area. This may reflect a lack of sensitivity (or even of validity) for this particular measure, or alternatively it may indicate a slightly greater use by disadvantaged parents of two of the language environment criteria, namely encouragement of nursery rhymes and the telling of stories; in a number of interviews these behaviours were advanced by parents as alternatives to what they saw as the more difficult task of reading to their children.

On both measures of meeting attendance - for the reading and mathematics programmes - mean scores of parents from the advantaged area were well ahead, by 35 and 13 per cent respectively. This particular comparison does not discriminate between working and non-working parents and is therefore biased in favour of the advantaged mothers, fewer of whom went out to work than did the disadvantaged mothers. But it is worth noting that four working parents in the advantaged area took time off from work to attend the meetings, compared to an estimated two working parent 'attenders' in the rest of the sample. (These exceptions clearly formed part of the experimental group, rather than the working parents group, as discussed in the next sub-section.) Overall, the meeting score comparisons reflect both the relative opportunity to attend as well as mean attendance levels for all parents in the sub-samples. Further sub-division to analyse the comparisons in more detail could not be statistically justified.

#### 6.424 Sub-samples from black and other groups

The question of whether comparisons should be made between the academic and other performances of Black and Non-Black children is an extremely sensitive issue, open to considerable misunderstanding. Jensen (1969), for example, wrote a much criticised paper (e.g. Kagan, 1969, Richardson and Spears, 1972) in which he analysed the conflicting data on the contribution of heredity and environment to the lower cognitive performance of American Blacks. More recently (1977) he has somewhat modified this position, on the basis of new research which highlights the environmental factors operating in this area.

No position is taken on this issue in the present study. However the comparison between the scores of children from advantaged and disadvantaged areas, cited in the previous sub-section, suggests that whatever the proportions of heredity and environment, and of interaction and other effects, the influence of environment on the performance of the 4-year-old child is massive. Attempts to blur or ignore the data on the situation of Black children, particularly those in areas of disadvantage, are a way of ignoring the gross deprivation which history and modern society has imposed on these children. The answer to their problems is not to avoid citing such data in a research context, but to see it

as an indication of the level of social and educational need which has yet to be ameliorated.

The present comparison looks at parent behaviours as well as the children's performance. Seven parent behaviours and two criteria for attendance at parent group meetings were used for the parent comparison. Eight performance variables were also selected, covering the children's attainment and ability scores in the nursery battery.

Children of West Indian and African parents (only a few were Africans) were grouped in one sub-sample - these were categories B and C (2 and 3) in the ethnic groupings derived from the parent interviews; all the remaining children - mainly English, but with smaller groupings of children of Irish, Asian and other descent - were placed in the other sub-sample. To ensure a fair comparison, the scores of parents and children at the one school in the advantaged area were excluded from this particular analysis (the scores of the two Black parents and children in the nursery sample at that school were likewise excluded). The scores are thus all derived from the parents and children at the five schools in disadvantaged areas (48 Black, 87 Non-Black).

The examination of the nursery test performance of the Black children on each of seven key variables showed virtually no difference between the Black and Non-Black performance levels. On three of the variables the mean performance was above that of the Non-Blacks: Bender Gestalt 4 p.c. (prob. .57); Infant Reading Test 1 p.c. (.91); and WPPSI Information 1 p.c. (.86). On four variables the Whites showed higher means: E.P.V.T. 3 p.c. (.69); Mathematics test 6 p.c. (.33); Reading awareness 6 p.c. (.47); and WPPSI Block Design 2 p.c. (.80). These scores are cited in some detail (full results in Appendix D 9) to emphasise their similarity, in sharp distinction with the noticeable and fairly consistent differences between groups whose performance is normally expected to differ - between boys and girls, and between advantaged and disadvantaged schools. The fact that the tests showed the expected differences on the two latter comparisons suggests a reasonable level of construct validity for the battery as a whole, and adds emphasis to the tentative conclusion that Black and White children within this disadvantaged sample do not in fact differ overall in early attainment and ability.

On the eighth test variable in this selection, Black children scored far above the mean of the rest of the sample. An ability to copy a rhythmic tapping pattern, which was shown in the author's previous research (Barker, 1976) to correlate .44 with later reading performance, yielded a Black mean score of 5.06 in the nursery battery, compared with the remaining sample's mean of 2.77, and a mean of 6.44 in the mid-test battery, compared with the remaining sample mean of 4.27. t-tests of the differences yielded the unusually high figures of

5.28 and 4.42. This particular skill was measured across all schools. The superior ability of Black children in the rhythmic tapping test was noted in the previous study as well. Its implications have yet to be examined in more detail.

In contrast with the evidence on the children's test scores, the means and standard deviations of the parent scores, as presented in Appendix D 9, showed fairly considerable differences on some of the key indicators. The 48 homes of Black or mixed race children showed average reading behaviour and language environment scores 10 and 14 per cent respectively below those of the remaining 87 homes in the disadvantaged sample, maths behaviour scores 19 per cent below, and television viewing hours (scores being a negative function of viewing hours) 18 per cent above those of the other homes. On the other hand Black parents expressed a 10 per cent more favourable reading attitude, while the television 'behaviour' (control over the child's viewing) and degree of parent-child cooperation were the same as those of the remainder of the sample. The average meeting attendance scores were calculated for all the parents in disadvantaged areas who were not working - these figures indicate thus the degree to which parents of the particular sub-sample chose to participate in the programme, in terms of their apparent freedom to attend or not. Here again the Black parents showed mean attendance levels 36 per cent below those of the remaining disadvantaged sample for the reading programmes, and 29 per cent below for the mathematics programmes. The probabilities of the differences ranged as low as .001, with only two differences favouring Black parents.

(The high probability figures for meeting differences are in reality a methodological artefact, owing to the method of averaging meeting attendance scores and calculating variances on the basis of the whole of a sub-sample, rather than over the parents who attended a particular set of programme meetings. This was necessary in order to arrive at sub-sample comparability; to have selected 'attending' parents within a programme group and compared their attendance scores would not have reflected the overall participation level within a sub-sample. The disadvantage of this method is that the large number of zero scores inflates the variance and reduces the apparent importance of the finding. Naturally this problem does not arise with any of the other parent or child variables, on which every individual is scored and the ordinary probability figures can be determined.)

There are several points that need to be emphasised in this comparison of ethnic parent behaviours. The behaviours have been scored in terms of European cultural preconceptions as to what is academically desirable in a home environment. The variables used do not in any way (with the exception of parent-child cooperation) reflect the warm and caring environment which many Black



parents provide for their children, and still less do they reflect the grave disadvantage that these parents face in terms of race prejudice and discrimination at the hands of some members of the host society. The sizes of the standard deviations also need to be studied, as reported in Appendix D 9. They show a considerable overlap, indicating that there are many Black homes with behavioural scores well above those of many White homes. There is thus no homogeneous pattern of low Black scores, but rather a wide range with the lower means perhaps reflecting the greater level of social need and poverty among the Black community. It needs to be emphasised that on almost every variable some Black homes were to be found among the handful of top scorers. One of the most cultured homes visited by E, as judged by the quality of the conversation during the interview and the deep awareness shown of the social and economic factors influencing British society today, was a mixed race home with a Black father and a White mother.

In the light of the finding that there is little or no difference on the nursery test performances, it may be asked why it has been found in so many studies (Little, 1978, is only the latest example) that West Indian children in English schools often have noticeably lower levels of educational attainment than do their counterparts in the host community? It has been claimed by some writers that this is due to prejudiced teachers or to an educational system that is not geared to the West Indian cultural ethos. However, the evidence on the differing ethnic levels of parent behaviours in this sample suggests that it could well be the lower level of home commitment that is responsible, to some degree, for the relative failure of Black children to capitalise on the educational experience provided for all children. This would be in line with the finding, reported frequently in the literature, that the parents' behaviour patterns or 'home environment' have a major impact on children's later attainment (Vernon 1969, Karnes 1973, White and Watts 1973, and Bronfenbrenner 1974 c, among many others). The issue has been discussed at some length in the introductory sections.

The numbers of Asian parents in this study were insufficient to enable a statistical comparison to be made between them and the White parents, but in the Asian homes visited the author did gain an overall impression of deep commitment to educational attainment and to the home 'work' environment necessary for such attainment. This, again, is borne out by the relatively good performance of Asian children at school, as reported by teachers in the sample areas, thus raising serious doubts about the level of claimed teacher prejudice or hostile cultural ethos, since those failings, if present, should also affect Asian performance.

The question of why the parents' 'home educational' behaviours differ between the groups is not one that can be answered in any degree within the

present study. Cultural concepts, attitudes and patterns of behaviour are undoubtedly part of one explanation. Another reason, suggested by an infant teacher, is that apart from the extreme social disadvantage which so many Black parents face, there is also a history of several centuries of slavery and lesser forms of serfdom in which the parents traditionally came to believe that all the educational power lay in the hands of an authority in which they had no share. While this attitude is also prevalent among disadvantaged White parents, the average Black parent may be even more inclined to leave early education, in its widest sense, completely in the hands of the schools.

6.43 Comparison of .....

6.43 Comparison of parent group characteristics

The following section (6.44) looks at the design problems which arise from the degree of self-selection which was inevitable in the setting up of the various experimental and comparison groups. The present sub-section simply examines these groups, in terms of the means and the probability of the differences between the means. In contrast with the previous section, where comparisons of sub-samples were based on sample sizes existing at the time the data were coded, all the following comparisons of programme groups are based on the final sample of 159 surviving at July 1978. This is clearly necessary if the effectiveness of parent programmes is to be examined in detail.

Sub-section 4.226 and section 6.22 offer descriptions of how the parents were allocated to the various groups which are defined below. In brief, these groups can be partitioned in five different ways. For clarity, the five forms of partitioning are presented in a single table overleaf.

First partitioning..../

Table 20. Parent groups: partitioning

First partitioning

Parents accepting invitation to participate in a programme:	125	A
Accepters who started, attending one or more meetings	99	B
Accepters, non-starters who attended no meetings	26	C
Parents turning down invitation to participate in a programme:	34	D
Parents who said they were interested, but were working	31	E
Refusers who indicated that they were not interested	3	F
Total number of parents interviewed, surviving July 1978	159	A + D

Second partitioning

Disadvantaged area parents:	130
Working group	28
Non-working groups (excluding 1 refuser)	101
	<hr/> 129
Advantaged area parents:	29
Working group	3
Non-working groups (excluding 2 refusers)	24
	<hr/> 27

Third partitioning

Disadvantaged areas, parents attending meetings	80	(disad. experimental gp)
Disadvantaged areas, non-attending parents	21	(disad. non-attenders)
Disadvantaged areas, working parents	28	(disad. working gp)
	<hr/>	
Total	129	

Fourth partitioning

Disadvantaged areas, experimental attenders:		
Reading group allocation	48	(disad. reading group)
Mathematics group allocation	32	(disad. mathematics group)
	<hr/>	
Total	80	

Fifth partitioning

All schools, experimental attenders:	
Fortnightly meeting groups	83
Six-weekly meeting groups	16
	<hr/>
Total	99

The first partitioning is of interest in showing how the sample divided during the initial phase of interviewing. However, there is no particular value in analysing all these sub-groups in relation to the research design. It is the remaining five divisions (which in any case incorporate the groups from the first partitioning) that are of major interest.

The second partitioning presents a comparison between the 129 parents at the five schools in disadvantaged areas and the 27 parents at the single sample school within an advantaged area. In both cases the refusers are excluded as atypical parents in terms of the research objectives. It could be thought that this division indicates a smaller proportion of advantaged parents going out to work during a period when their children were not yet in full-time education. In fact, several advantaged parents who were in employment took time off to attend programme meetings. Close on 30 per cent of the parents in both social groups were employed, most of them in full-time jobs.

Specific comparisons between the mean parent behaviours and child nursery performance of the dyads in the advantaged and disadvantaged area schools have already been given in the previous sub-section and in appendix D 8. The comparison between these two groups is of particular importance, however, and a number of the analyses later in this chapter will focus on the differences.

The third and fourth partitionings refer only to the parents from schools in disadvantaged areas. There are various reasons why most of the analyses are to be based only on these parents. The original research goal was to study whether parent programmes would be effective in raising the early academic performance of children in disadvantaged areas. A sample school was added from an advantaged area because of the interesting comparisons it might afford. Some analyses will thus be based on the total sample and on specific comparisons between advantaged and disadvantaged sub-samples, but the disadvantaged groups will be the main focus of attention.

There are further statistical reasons for restricting most of the comparisons to the latter groups. A total sample of 5 disadvantaged schools and 1 advantaged school is not representative of the national or even the metropolitan school population, whereas the 5 randomly selected disadvantaged schools form a relatively homogenous sample, representative of an inner city's disadvantaged areas. The proportions and relative performance of the various programme groups and the corresponding children differ considerably between the advantaged and disadvantaged areas; the combining of these data within a single sample analysis not only blurs the pattern of the disadvantaged groups but also obliterates the differences between the two social groupings.

The third partitioning reflects one of the principal forms of comparison within the sample design, namely the three-way division of groups across the

disadvantaged sample - parents who agreed to participate and did attend at least one meeting; those parents who agreed, but attended no meetings; and those who said they could not attend meetings because they worked. The one 'refuser' in the disadvantaged areas is not included in these groups.

The fourth partitioning divides the 80 attenders in the disadvantaged areas into the randomly allocated reading and mathematics groups. This provides another basic form of experimental comparison. The evidence that there are one and a half times as many parents in the reading groups as in the mathematics groups is an artefact of the procedure for creating programme groups at any one school, and does not call into doubt the randomness of the parent allocation. When an equal number of each kind of programme group (reading and mathematics) had been formed at a school and there were only sufficient parents to form one further group, the decision was usually taken to set up an additional reading group, since the reading programmes were the main focus of interest. The 'selection' of this final group of parents was however rooted in the same chance geographical factors that were responsible for the sequence of most visits within a school interviewing schedule (as described in section 4.2226).

The final partitioning divides the 99 attenders, from all schools, into those attending the 2-weekly and those at the 6-weekly group meetings. This division was also randomly made when the groups were formed.

Appendix D9 presents five comparisons between the various programme and non-programme groups, based on eight key attainment and cognitive variables measured on children in the nursery (pre-test) battery, and on five main parent behaviours recorded in the interviews. It should be noted that for these group comparisons, unlike the population sub-sample comparisons in 6.42, the nursery mathematics score has been divided into its two sub-scores of 'numeracy' and 'mathematical concepts'.

As there were no prior hypotheses about the differences between the groups in four out of the five comparisons dealt with below, the probability statistics used for these comparisons are two-tailed. The deeper implications of the findings from the comparisons will not be discussed in the present sub-section; the whole framework of the experimental design and analysis will instead be reviewed in the following sub-section.

6.431 Disadvantaged areas: working parents vs. non-working groups

In essence, this comparison looks at the differences between the 28 parents who went out to work and all 101 parents who remained at home while their children were still at the pre-school age. On the whole, as shown in Appendix D 9, the working group parents and their children have a somewhat higher level of child nursery performance (the two-tailed probability of the differences reaching 0.03 in the case of Infant Reading Test and 0.06 for the WPPSI Block Design; other p figures range from 0.11 to 0.64). The two variables against the trend (Mathematical concepts and WPPSI Information) have p figures of only 0.83 and 0.74. On the other hand, parent behaviours show little or no difference between the groups, with p figures apparently at random level, apart from Television time and Television (control) behaviours where differences are in opposite directions with p figures of 0.38 and 0.31 respectively.

In so far as a series of differences across a number of variables shows a fairly consistent trend, it is possible to argue that even this moderate range of probability figures does indicate that the children of the two groups differ quite considerably, especially in their early reading level, in reading awareness and in the non-verbal cognitive skill of Block design. On the other hand the parent behaviours do not differ appreciably. It may well be that working parents had a higher level of behaviours at an earlier stage, before they went out to work, and it was that which contributed to higher performance in their children; alternatively the sub-cultural environment of the homes may on balance be more sophisticated than that of the non-working parents.

6.432 Disadvantaged areas: working parents vs. programme attenders

Since 80 of the 101 non-working parents attended programme meetings it could be expected that this comparison, between the 28 working parents and the 80 non-working programme attenders, would not be too different from the previous one. The differences are all in the same direction as in the first comparison, but with only a few exceptions these differences are reduced quite considerably. Probability figures for the set of variables favouring the working group children now range from 0.05 to 0.67, with only two variables still against the trend, as before. Parent behaviours remain at almost random levels of differences between the groups.

6.433 .... /

6.433 Disadvantaged areas: programme attenders vs. non-attenders

These comparisons are within the group of non-working parents, between the 80 attenders and the 21 non-attenders. With one exception among the 13 child and parent variables, all the differences analysed in Appendix D 9 favour the parents who attended one or more meetings, as against those parents who had agreed to attend but were not present at any meetings. The lowest probability of these differences was 0.31, on parent reading behaviours. Other p figures ranged from 0.27 to 0.92. The only difference contrary to this trend, Television behaviour, showed a p of 0.91.

The evidence about these differences is not only consistent but is also in accord with expectations. It can be argued that many of the parents who failed to attend any meetings would be likely to be those who home 'educational environment' and awareness of their educational role were lower than the characteristics of those who attended programme meetings. This is of course based on the cultural assumption that parents have an educational role in addition to their role as child-rearing caretakers.

6.434 Disadvantaged areas: reading..../



#### 6.434 Disadvantaged areas: reading attenders vs. mathematics attenders

The comparisons again show an almost totally consistent pattern favouring one group, in this case the 32 mathematics attenders, against the other group, the 48 reading attenders, over the 13 child and parent variables used here. The p figures range from 0.14 to 0.93, again covering most of the scale in between. The only exceptions to the trend are E.P.V.T. (p 0.94) and Television time (0.68).

While there are conceptual grounds for arguing that the differences between the groups in the first three comparisons are logical and justifiable in terms of the nature of the parents who would 'select themselves' into groups such as working parents, programme attenders or non-attenders, the fairly consistent differences between the reading and mathematics groups could suggest a serious lack of randomness in the principles followed by E in allocating parents into groups. The procedure for randomising the allocation has already been described (sub-sections 4.226 and 6.22). The possibility was considered that since the 'extra' groups formed late in the interviewing phase at any one school were usually reading groups, for the reasons given earlier, it may have occurred that a few parents who were more difficult to find at home and had therefore been interviewed later, had biased the reading group's mean scores downwards. This argument is countered by the fact that the 6-weekly groups, who were in several cases the last group to be formed at a school, showed no difference from the 2-weekly groups (see following sub-section).

One answer to this problem was given by the results of a comparison, over all schools, between parents randomly allocated to reading groups and those allocated to mathematics groups, whether or not they attended any meetings. These comparisons, on seven parent behaviours, show a random series of small differences favouring one or the other group. The probabilities of the differences were as follows - comparing 76 'reading accepters' with 53 'mathematics accepters' and citing the group with the higher mean in each case:

Reading behaviours	Reading group	0.50
Language environment	Mathematics group	0.88
Parent reading attitude	Mathematics group	0.42
Mathematical behaviours	Reading group	0.33
Parent-child cooperation	Reading group	0.08
Television time	Mathematics group	0.91
Television behaviours	Reading group	0.86

Evidence on attendance at programme group meetings, to be cited later in this study, offers a possible reason for the consistent difference between the sub-samples of reading and mathematics attenders in the disadvantaged areas.

The mean attendance figure for individual parents in reading groups was considerably higher than the mean attendance for parents in mathematics groups. The idea of attending a mathematics programme seemed strange to a number of parents, according to what they told E, whereas no parent questioned the relevance of reading programmes per se. The possibility was therefore considered that the characteristics of the mathematics non-starters were different from those of the reading non-starters. There were 13 mathematics non-starters (out of 53 accepters) compared to 13 reading non-starters (out of 76 accepters), but in other respects the comparisons provided a somewhat mixed pattern of differences. Even if only the disadvantaged prior dropout sample is examined the results offer no clear pattern of differences.

The difficulty of interpretation when using this kind of statistical hair-splitting, and preferable methods of handling the analysis, are discussed in the next main sub-section, 6.44. Related issues have previously been discussed in sub-section 5.15, on the significance concept.

6.435 All schools: 2-weekly reading attenders vs. 6-weekly reading attenders

As there was only one mathematics 6-weekly programme group it was decided to describe here only the comparison between the 47 attenders in the 2-weekly reading groups and the 13 attenders in the 6-weekly reading groups. With the exception of Language environment ( $p$  of 0.07 in favour of the 2-weekly groups) and Television time (0.01 in favour of the 6-weekly groups), the results showed an apparently random set of probability figures for the differences favouring one or the other group. The evidence listed in the last table of Appendix D 9 suggests therefore a successful randomisation procedure with no indications of selective prior dropout favouring the 2-weekly or 6-weekly attenders.

#### 6.44 The experimental design and analysis

The introductory note on the intervention project (section 4.10) discussed the dependence of the experimental design on the form of the planned analysis as well as on the expected characteristics of the sample groups. This section reviews the question of design in some detail, since the adequacy or otherwise of a design is one of the cardinal features in judging any results that might be obtained. The nature of the planned analysis is also reconsidered.

Perhaps the most troubling factor in the present design is that parents with better behaviours (in terms of the interview categories) were also the parents most likely to take part in the experimental programmes. This factor is found in virtually every intervention study; even among a strata of disadvantaged parents it is those who are either rather less disadvantaged or more 'upwardly mobile' who tend to make greater use of programmes offered to them and who are less likely to drop out of the programmes. It is the same factor which makes control group comparisons so difficult, since it is only a genuine invitation to participate which can reveal the degree of take up within an experimental group. On the other hand such invitations cannot normally be made to those deemed to form part of a control group; ethical problems arise if the invitation is not made in good faith.

The presence of this 'invitation bias factor' introduces a design complexity. The most basic design method requires the prior random allocation of subjects into experimental and control groups (design 4 in Campbell and Stanley, 1963). This division is, however, confused if experimental treatment is dependent on both the acceptance of an invitation to participate and the attendance at or taking up of the treatment. If only those accepting and participating in a programme are regarded as the experimental group, and all those randomly allocated to a control group are regarded as the controls, any measurable effects of intervention are erroneously strengthened - and thus invalid. Even if one were to expand the experimental group to include all those accepting the invitation, whether or not they eventually participated, one still has a sample biased in favour of the experimentals if the control group includes a range of both accepters and refusers. On the other hand, if one compares an entire group of randomly allocated experimentals, including accepters and refusers, with a similar random control group, then one may be seriously diluting the measured effects of the intervention programme on those who actually participated; nevertheless, the latter procedure cannot be faulted as invalid. Campbell and Stanley (p.16, 1963 reprint) describe a whole range of methodological problems associated with this latter method which they term 'invited remedial treatment'.

One of the alternative designs that were considered at an early stage in

the project was to offer all parents the opportunity to participate; those who said they were unable to do so because they were working, and those who simply refused the invitation, would both be rejected from the sample. The remaining sample would be randomly divided into experimentals and controls. A basic criticism of this alternative is that both the experimental and control parents would be different from the ordinary population from which the original sample was drawn, since both groups would consist of the kind of people who accept (and who are free to accept) an invitation to participate. Thus even though ostensibly based on a true experimental design, the results would suffer from what Campbell and Stanley describe as the interaction of selection bias and the experimental variable.

It was clearly preferable to look more closely at the possibilities of Campbell and Stanley's 'true experimental' design 4, namely the prior random allocation of all parents into two groups, experimentals and controls, followed by invitations only to experimental parents and the provision of programmes only for those parents who accepted the invitation. Thus the whole group of experimentals would be compared directly with the controls, without any exclusions.

Apart from the unavoidable dilution effect previously referred to, there are a number of objections which are to some degree related to the particular circumstances and planned analysis of this experiment, or quasi-experiment. Assuming the need for a reasonably large sample, spread over a number of schools, the possibility of random allocation of the schools into experimental and control samples offered a solution that could eliminate some of the objections listed below. However, with the considerable variability in the degree of disadvantage found in the catchment areas of the different schools, it is unlikely that any adequate randomness could be found with less than five schools in each sample of schools - thus at least 10 schools would be needed. A venture of such a size would not have been feasible for a research project involving only one worker.

A crucial objection to a design requiring an initial random allocation of parents into experimentals and controls is based on numerical considerations. It was expected, and found in practice, that a considerable number of parents dropped out during the programmes; one-fifth of those who initially agreed to participate never turned up to any meetings. There were also continuing attrition losses. The reality of the project confirmed the expectations of these considerable losses. Yet the kinds of analysis planned for the research required a reasonable size of sample - certainly well above 100 after taking account of losses.

Briefly, the initial sample was 204; 45 were lost before the post-tests. Of the remaining 159, 31 were grouped by E into the 'working' group - they had offered valid reasons such as being employed away from home or (occasionally) as unable to leave home because they were working as child-minders or had a large

number of young children, or were employed in a family shop on the home premises. Only three refused the invitation because of an admitted lack of interest. In view of the small number of absolute refusals it was considered legitimate to eliminate these three cases from any of the comparisons between experimental and other sub-samples, although the three were included in the 'whole sample' analyses. However, even the division into working and non-working groups could not be absolute. While most working parents responded to the invitation to attend programme meetings by saying 'I'm sorry, I'd really like to attend, but I work and wouldn't be able to get away', there were a small number who were able to take time off to attend and were thus included in the experimental group.

In total, some 125 parents both accepted an invitation and were free or could find the time to participate in the programmes. Had the full sub-sample at each school been divided initially into two random groups, this would have left approximately half that number of experimental parents - say 62. If only the schools in disadvantaged areas are considered, the experimental and control groups for the bulk of the analyses (on the disadvantaged) would have been reduced to 50 each.

While this alternative design would have offered the two major advantages of sample equality and the format of a true experiment, it would also have brought a number of other problems apart from the reduced size of the experimental sample.

With the additional loss of 22 per cent who accepted the invitation but never attended any meetings, this would have meant (in the randomly divided sample) only an estimated 40 parents filling between one and three groups at each of the five schools in disadvantaged areas. There would have been additional complications in allocating individual groups to reading or mathematics programmes (or eliminating the mathematics programme altogether) and in providing separate groups for morning and afternoon parents. In short, the problems which arose in the present design over the smallness of group numbers and over group randomisation would have been compounded in an experimental sample only half that size.

The importance of numbers has been discussed in the previous chapter, when dealing with the choice of statistical methods. It is not in any way related to the trivial and obviously misleading strategy of raising sample numbers to ensure the 'significance' of simple comparisons. Rather it is concerned with the basic tenet that a model which takes into account several dozen contributory factors, each of them assumed to contribute to ultimate reading and mathematics attainment, requires the kind of sophisticated analysis that becomes increasingly difficult as sample sizes become smaller.

The randomization into experimental and control groups may have brought other complications as well. With half the parents excluded and aware that they had been excluded, a number of these controls may have been motivated to borrow materials and ideas from those friends who were attending group meetings, creating a 'contamination' effect that would have been impossible to measure or even assess to any degree. (There was in fact some minor admitted 'contamination' between parents in the reading and mathematics groups at two of the schools.) There might also have been a kind of Hawthorne effect among the participants, aware that they had somehow been specially chosen for the programmes; this could have inflated the effects of the programmes beyond what might have been expected normally.

The considerations which led to the decision not to divide the sample randomly into experimental and control groups have been outlined in the past few pages. To what extent do the existing project groups - particularly the working parents vs. programme attenders and the reading attenders vs. mathematics attenders - provide adequate and reasonably similar comparison groups?

The statistical comparisons presented in the previous sub-section (6.44) yield the answer to this question. Early hopes that the working and programme attender groups would be reasonably comparable were negated by the finding of fairly consistent pre-test differences, including a crucial  $p < 0.05$  difference in favour of the working parents on the pre-test scores of the Infant Reading Test. The surprising pre-test differences between the randomly allocated reading and mathematics groups, when only the disadvantaged attenders are compared, have already been discussed.

The multivariate nature of the sample suggested one further attempt to assess whether the different parent groups might be seen as comparable. Discriminant analysis could have shown whether or not the groups differed 'significantly' and, provided the results did not yield F statistics below a probability of 0.05, it could have been argued that the programme attenders could be compared with the working parents as a pseudo control group, or alternatively that the reading and mathematics attenders could serve as mutual control groups. Some discriminant analyses were carried out, but it became evident that this approach would involve several major limitations. Firstly, a large number of pre-test variables would have to be included in the statistical comparison; secondly, several basic assumptions would have to be made on what was acceptable - for example, that F figures above 0.05 could be disregarded as not indicating important differences between groups; and thirdly, that subsequent analyses would have to be restricted to variables covered by this formal certificate of equality. There seemed thus little point in developing this approach further.

While it had been intended from the outset to carry out both elemental

and advanced forms of analysis, the mixed results from the parent group comparisons, as well as the evidence of the widely differing nature of the advantaged and disadvantaged sub-samples, compelled a certain reorientation of the planned strategy of analysis.

To sum up the main problems facing the study at this point:

- a. Not only did the advantaged and disadvantaged sub-samples differ considerably (as could be expected) on child nursery performance and parent behaviours but there was also the difference that whereas in the advantaged sample it was the parents who stayed at home with their pre-school children who showed superior scores on the project variables compared to the (relatively few) parents who went out to work, in the disadvantaged sample the reverse obtained, with the children of working parents showing considerably higher scores.
- b. The problem of identifying equivalent comparison groups for the programme attenders has been described at some length. The use of working parents or non-attenders as comparison groups (in the strict statistical sense of measuring the probabilities of the post-test differences) could bias the results quite seriously, in one or other direction. Even within a particular comparison there would be possibilities for selective focusing on variables which tended to support a particular experimental hypothesis.
- c. The random allocation procedures for reading and mathematics groups appeared to achieve randomly differing results for the initial allocation across the whole sample, but yielded a consistent trend in favour of the mathematics groups if measured in terms of those who attended programme meetings in the disadvantaged areas. Should a comparison then be made between those allocated to reading and mathematics groups respectively, over all schools, advantaged and disadvantaged? How should the fact of a major difference in performance of the advantaged mathematics group and the disadvantaged mathematics groups be interpreted within such an analysis?

These problems are not new in intervention studies and much ingenuity has been shown by other researchers in the ways found to minimise their effects. But ingenuity and honesty in the tackling and presentation of hypotheses and analyses of differences are not enough to cope with the vagaries of fortuitously large initial differences in randomly allocated samples of only moderate size, especially if techniques for the statistical removal of outliers are not to be adopted. There are good grounds for doing that removal in certain circumstances but it can also be misleading in a situation of testing intervention hypotheses on a sample from a population.

The finding of the many differences between the parent groups, overlaid by the population sub-sample differences described in section 6.43, presented

thus the choice between two alternatives: either to continue with the variety of analyses as originally planned, endeavouring to account for specific sub-sample and parent group differences in post-test attainment and to assess these in multivariate analyses of variance and covariance, as well as carrying out two or three path analyses; or alternatively to examine the data using the three forms of analysis described below:

- i. A brief review of correlations across the whole sample and a separate review across the disadvantaged sample, looking for suggestive pointers but aware of the grave statistical limitations of any findings from this form of study.
- ii. A non-statistical examination of pre-test and post-test group scores on important attainment variables, without attempting to interpret the findings beyond the limits of a brief examination of the score pattern.
- iii. A considerable expansion of the planned path analyses; the interrelationships between the different conceptual clusters of variables would be examined in separate but equivalent path models for each of the identifiable parent groups and sub-samples within the project.

This second alternative appeared preferable, especially as it was possible to use the path analyses to develop comprehensive and interpretive comparisons between the various groups and sub-samples in terms of the probabilities and sizes of the path coefficients. These parameters would serve as a measure of the contributions made by child, parent, programme and other variables to the variance of the post-test attainment criteria. It was decided to carry out 10 such analyses, to obtain a clearer picture of the sample and its development over the project period than could have been obtained with the first alternative.

The statistical rationale for path analysis has been presented in chapter 5.

In conclusion, it can be argued that it is a legitimate strategy to allow the nature of the initial data structure, as it reveals itself in a close examination of the pre-test results, to suggest modifications in the preferred methods of analysis.



6.50      Data: reliability and validity

The theoretical foundations on which reliability and validity are assessed in this study have been described in sections 5.14 and 5.15. It is apparent that even within the boundaries defined in those sections there will be many points at which subjective judgements will have to be made as to the retention or rejection of a design variable. At the same time the evidence on the characteristics of these variables is crucial to the interpretation and conclusions to be drawn from the results presented in the final three sections of this chapter.

A fundamental difficulty with any approach to reliability and validity is that to some degree the two characteristics are negatively related. House et al (1978) describe the problems which arise when major evaluation studies are based only on tests with proven high reliabilities; the over-emphasis on objectivity comes close to the concept of operationalism, where everything must be jointly seen and specified in order to be considered 'true', even though the instruments may be invalid for the purpose for which they are used.

At the other end are what might be described as 'soft' measures which have an intuitive appeal, such as motivation, but which are difficult to assess in any reliable manner despite the conceptual arguments that such measures are more closely related to educational attainment, for example, than are 'hard' measures of perceptuo-motor skills.

Cannell and Kahn (1968) see measurement inadequacy in terms of three aspects - validity, reliability and precision. These characteristics are not wholly independent and do in fact have a complex interrelationship in which there is a trade-off between sensitivity of measurement or validity, and accuracy or reliability.

The ideal is to find measures which occupy a midway position between the hypothetical extremes in which either reliability or validity are so emphasised as to call into question the balance between the two. At such an intermediate point it is possible to take statistical and conceptual account of the limitations of this necessary compromise.

Section 6.51 presents the evidence on the reliability of the variables used in the study.

Section 6.52 focuses on a particular aspect of validity. As explained in the theoretical discussion of this concept (section 5.15), the content validity has already been defined in the description of the tests and interview instruments. The predictive validity, seen as the contribution of the variables to the total analysis, will become evident in the final three sections of this chapter, 6.60,

6.70 and 6.80. The coming section 6.52 deals mainly with the nomological validity of the variables, as defined in the theoretical discussion.

### 6.51 Reliability determination

The discussion in section 5.14 emphasised the importance attached to reliability estimates for the analysis of the data obtained in this study. It was because of this importance that more than four weeks were spent on reliability testing, out of a total of approximately 25 weeks taken for testing the children at various stages in the programme.

One factor which may have affected the overall reliability of the study, but to a very marginal extent, was the practice by E of providing the teachers with verbal feedback and the schools with a brief written report of a few lines on particular weak or strong points found in each of the children tested. This was done deliberately, as a matter of fairness to the children concerned and in response to the schools' kindness in providing facilities for the research. For example, one nursery child who could hardly speak and was considered extremely backward was found to have non-verbal intelligence scores well above 130, more than 50 points above his verbal scores. The contrast was confirmed later in the reception class tests. A neglect to mention such a finding could possibly delay the chance of capitalising on this strength. There were many other cases where the test findings helped both nursery and reception teachers, mostly, however, in the sense of confirming their own impressions gained through class work with the children.

It was only in a small number of cases, such as the one quoted above, where test findings came as a surprise, both to the teacher and to E. In general, therefore, the effect of the feedback of information on the total academic situation could only have been minimal. In these terms it is a small price to be paid for the advantages of research in an everyday school situation.

Appendix D6 lists the reliabilities of all the 56 variables used in the analysis.

#### 6.511 Pre-test reliabilities

For the nursery battery, it was decided to test and re-test one age cohort of children in the pilot school. Initially there were 25 children, but as with all testing of nursery children, prolonged absences and other factors reduced the available total to 21. It was expected that there would be a variety of problems in testing these children, some of whom frequently missed a day. It was also found that there was a considerable variation in mood and performance among the young children - for example, a child getting up from a jigsaw puzzle to come for testing would respond very differently from one who had just been involved in an altercation with a diminutive friend, and differently again from a child who had been told that a class birthday party was to be held later

Table 21. Pre-test reliability data (test-retest correlations)

Variable	Tests a/b	Tests b/c	Tests a/c	Test mean (root mean square)
Rhythmic tapping	.72	.64	.33	.58
M.F.F. visual	.61	.49	.24	.47
M.F.F. tactile	-.02	.42	-.13	-
Bender Gestalt	.74	.73	.58	.69
Self-picture test	.89	.86	.88	.87
Piagetian test	.72	.90	.71	.78
Maths test	.86	.87	.90	.88
Reading awareness	.41	.67	.50	.55
Infant Reading Test	.96	.97	.94	.96

that morning or afternoon. Accordingly it was decided to carry out the battery of tests three times and use as the test reliabilities the square root of the mean of the three squared test-retest correlations, i.e.:

$$r_{tt \text{ ave}} = (r_{tt1}^2 + r_{tt2}^2 + r_{tt3}^2)^{\frac{1}{2}}$$

The root mean square is clearly a more appropriate statistic for estimating an average correlation than would be a simple average, since correlations are not additive whereas variances are.

The problems in ensuring repeat testing of these children proved greater than expected. Absences, the impending departure of a few children on holiday, and other factors meant that test-retest intervals varied between a few days and six weeks. While four of the 25 children had to be sacrificed since only one test was available on them, of the 21 remaining, 20 were tested on all three occasions and one on two occasions.

The results appear in table 21. Nine of the 14 nursery tests were used in the reliability assessment. Of the five not used, four were the WPPSI sub-tests, where it was decided to use the reliabilities to be obtained in the later reception reliability assessment; the fifth was the E.P.V.T., where it was decided to use the reliability figure given in the manual for this test.

(The E.P.V.T. figure is .88 for test 1, for the youngest age range for that test (5:0-5:11); it is given in the Brimer and Dunn (1962) manual. The Brimer and Dunn (1973) manual on the full-range test (3 to 18 years) covers only administration. However, other 1962 E.P.V.T. reliability data suggest little change over age.)

There is a lot of variation between the three test-retest correlations for the rhythmic tapping and MFF visual tests. The reliabilities themselves

are not too satisfactory. However, in the light of the fact that both tests had only a limited range of scores (MFF having been reduced to half the original planned total, as explained later), it is understandable that reliabilities would tend to be low. The MFF tactile reliability is clearly unacceptable. The MFF visual reliability (on 6 items) can be compared with the MFF split-half error reliability of .46 on 9-year-olds and .52 on 12-year-olds, obtained by Cairns (1977), and the Ault (1976) review of a number of studies, showing a test-retest error reliability range of .23 to .43, with the shortest period (3 weeks) yielding .39.

Bender Gestalt reliability is satisfactory in the light of a wide range of studies reviewed by Koppitz (1975), where figures range from .53 to .90, with an average of .74, samples ranging from 5-year-olds upwards. The reliability of the Piagetian test is quite acceptable - though probably biased upwards by the high proportion scoring zero or close to zero at nursery age. Reading awareness shows a rather low reliability, but as this test is based on answers to a series of 'talk' questions, with a variety of possible answers in each case, it is unlikely that this figure could be much higher than the one obtained here.

The self-picture and mathematics correlations show highly satisfactory reliabilities. Goodenough (1926) reports a series of studies by Terman Lewis showing draw-a-person reliabilities of between .80 and .90 within an unselected age group having MA's from 4 to 10.

Perhaps the most surprising figure is that of the Infant Reading Test, a test which was shown earlier to have an uncomfortably high standard deviation and positive skew; this has an exceptionally high reliability, as shown by all three inter-test correlations. In previous research by Barker (ibid), it was found that the test-retest correlation of I.R.T. was .934 on a small sample of 10 reception class children; an inter-observer reliability correlation, also on 10 children, gave a figure of .986.

One test which was included in the nursery reliability battery, but yielded almost no data here, was the distractibility assessment. It was a surprising coincidence that hardly any of the participating 21 children showed distractibility at a level where they could be scored. In fact, several of the four children whose absences meant that they could not be tested a second time were highly distractible. Rather than rely on a trivial correlation mainly based on zeros, it was decided to utilise the figure obtained in the reception reliability test for this characteristic.

## 6.512 Mid-test reliabilities

The reliability tests on the reception class children offered fewer problems of administration. The children were now in full-time and compulsory attendance, so that the only absences to be contended with were the genuine ones caused by illness. A total of 30 children, from three different schools (the pilot and two sample schools) were assessed on two occasions, in virtually every case over a strict one-week test-retest period. In no case were the same children used as had been used for the nursery test-retest assessment.

The results are given in table 22. None of the figures are so low as to cause concern. The distractibility correlation is reasonable in the circumstance that it relies on fairly subjective estimates of test behaviour.

There are some interesting contrasts over the five tests whose reliabilities were also estimated in the nursery. It could be expected that, with the increasing maturity and stability of the children, reception reliabilities would be higher, and this is indeed the case with four of the tests - rhythmic tapping, MFF visual, Bender Gestalt and Piagetian. (The doubling in the length of the MFF no doubt also contributes to the increased figure here.) An exception is the self-picture test, which shows a lower reliability.

The explanation for this could be that as children grow older there is an increasing element of artistic creativity in their drawings - there were in fact many mid-test attempts to portray hair styles and particular items of clothing, alongside very primitive representations of the human form. Judgement of the number and quality of details in the drawings will clearly diminish in reliability with the intrusion of these creative effects.

A comparison of the WPPSI test-retest figures can be found in the American WPPSI manual (Wechsler, 1967). This offers test-retest data, ranging from 48 to 117 days, of .75 for Information, .60 for Sentences, .93 for Picture Completion and .76 for Block Design. Split-half reliabilities, corrected by the Spearman-Brown formula, are .77 and .83 for Information and Sentences, and .81 and .83 for Picture Completion and Block Design. (These figures were based on 5-year-olds.)

Table 22 . Mid-test reliability data (test-retest correlations)

Variable		Variable	
Rhythmic tapping	.63	Distractibility	.61
MFF errors	.59	WPPSI Information	.97
Bender Gestalt	.85	WPPSI Sentences	.91
Self-picture test	.73	WPPSI Picture Completn	.83
Piagetian test	.84	WPPSI Block Design	.87

### 6.513 Post-test reliabilities

The final battery of five tests was likewise assessed for reliability. Here 35 children were used, taken from three sample schools for a one-week test-retest. Again care was taken that no children were included who had been used for any of the previous test-retest assessments. The results appear in table 23.

One of the children assessed in this battery showed how seriously a child's own moods can contribute to unreliability. On the first occasion he was playful to excess; on the retest occasion he had arrived at school in an angry mood after a dispute with his mother and was still in that mood when he was tested later. The discrepancies in scores for this child were relatively serious. However it was felt that in the circumstances the boy could be seen as a normal part of the sample and thus his test and retest scores were retained in the reliability measures.

Table 23. Post-test reliability data (test-retest correlations)

Variable		Variable	
Infant Reading Test	.97	Mathematics test (combined)	.81
Southgate Reading Test	.79	Piagetian test (one failure)	.73
Daniels and Diack R. Test	.987	Piagetian test (two failures)	.72

The test-retest results are reasonably satisfactory. The high figure for the I.R.T. is in line with the pre-test reliability statistic. The Southgate figure is on the low side, in comparison with the parallel form correlation of .95 quoted in the test manual (Southgate, 1959); however the latter correlation was based on a sample ranging from 5:8 to 8:1 years, where blind guessing problems would be less likely to arise. On the other hand the Daniels and Diack figure appears to be unwarrantably high; this figure is inflated by the fact that one-quarter of the reliability sample scored zero on the test on both occasions. The mathematics reliability (based on the combined concepts and numeracy scores) was rather low, especially when compared with the pre-test figure of .87. The Piagetian figures, based on the two systems of scoring described in Appendix A3, are in accord with expectation for a test of that nature, in which there is some uncertainty as to the child's grasp of the concepts involved.

Perhaps the most serious question in regard to the reliability of the post-tests was the fact that the E who had devised and administered the intervention programmes was also the person who administered these final criterion tests. This is an inevitable result of the fact that the research is carried out as an individual project and funds are not available to pay someone else for the many weeks of testing involved in the post-test measures. On the other hand it can be pointed out that E found it difficult to relate the names of the 159 children in the final sample to the particular names of the mothers who had participated

in programmes or who had been in the 'working group' more than a year previously; in many cases E had continued to meet both working and programme parents at the start or finish of a school day and the boundaries distinguishing the two groups had blurred. Only in possibly eight or ten of the post-test cases did E have a clear awareness that the children's mothers had been in programme groups, sometimes because the children themselves talked during the testing session of having been given 'reading games' by their mothers.

#### 6.514 Parent interview reliabilities

The estimation of acceptable reliability figures for these variables offered considerable problems. Section 5.141 describes the reasons why it was decided to take both observer reliability and the factual validity of the data into account in determining overall reliability figures for each of the parent interview variables. Some ten variables were derived from the interview protocol. One of these was ethnic group, which was clearly a categorical variable. Of the remaining nine, parent mathematics attitude and parent school attitude were eliminated at an early stage of the analysis when it was found that the responses were often almost random, suggesting that these were issues to which little thought had been given. At the time of the interviews the children were not attending formal school and the mixing of nursery and school-oriented items within the parent school attitude variable may also have negated the potential value of this item.

It was felt that it would be unwise to attempt to obtain a direct measure of observer reliability in view of the relatively sensitive nature of the questioning, in which parents were asked for example whether they read to their children, or exercised any control over the children's television viewing. Instead it was decided to ask some parents for permission to tape record the interview, for the expressed purpose of carrying out a reliability check on E's interview data. These parents were of necessity selected on the grounds of E's experience in the first few minutes of the interview, or (in some cases) on the basis of having become acquainted with the parent in the nursery setting. All the parents chosen were seemingly outgoing and willing to talk frankly about themselves. The taped interviews were given to a colleague so that she could listen to them and code a separate set of forms.

This was clearly not a completely satisfactory solution, as listening to each tape involved a considerable amount of work, with frequent interruptions to play back parts not heard clearly. The work required of the colleague, together with the difficulty of finding parents whom E could assume would be happy to be tape-recorded, meant that only six such interviews were ultimately recorded.

These interviews were used to arrive at an initial measure of inter-observer



reliability, based on the correlations between the scores for each composite variable as assessed separately by E and his colleague. The correlations were however open to amendment on several grounds, especially because of the inevitable discrepancies between the two methods of scoring the interviews and the small number of cases assessed.

Published data on interviewer reliability offered some useful parallels on which modification of the initial reliability figures could be determined. Figures bearing some relevance to the present study were the following:

Hyman et al (1954)

Re-interview by same interviewer on:	Level of education	86%	(r 0.93)
	Church attendance	89%	(r 0.94)
Re-interview by diff. interviewer on:	Level of education	80%	(r 0.89)
	Church attendance	67%	(r 0.82)

Cox and Butter (1976)

Inter-rater reliability of assessment of feelings      0.80 to 0.85

Yarrow et al (1970)

Inter-rater reliability in maternal interviews      0.70 to 0.90

On the basis of these figures the following adjustments were made to the initial estimates of rater reliability:

Table 24. Adjusted reliability values for parent interview variables

Variable	r between taped and actual interviews	Adjusted r
Reading behaviours (parent)	.73	.80
Language environment	.93	.95
Reading attitude	.96	.95
Mathematics behaviours	.84	.90
Parent-child cooperation	.50	.60
Television time (child viewing)	.79	.90
Television behaviours (control)	.52	.60

Relatively higher adjustments were made to the reliability estimates for parent-child cooperation, television viewing and the various behaviours on the grounds that all these variables were based partly on parental report and partly on E's observation of the home environment and parent-child relationships at the time of the interview; it was difficult to assess these fully from a tape recording of the interviews. It should also be stressed that the only purpose of the adjustments is to ensure a more widely based set of figures for the disattenuation process, using other evidence on this kind of interview reliability alongside the

evidence on the comparison between the tape-recorded and the actual interview codings. Clearly the latter would be likely to differ more than would inter-rater comparisons at the same interviews.

The other major aspect of the reliability of these measures concerns the factual validity as defined in the earlier theoretical discussion. With the time constraints on E it was again necessary to look to outside sources for comparable data on the factual validity of various reported behaviours. The list below cites some of the very limited number of studies which have examined this issue:

Per cent invalid statements

Cannell and Kahn (1968)

Contributions to the Community Chest	40
Presence of heart disease in the family	40
Presence of mental disease in the family	75
Presence of asthma in the family	16 to 29

Blum and Naylor (1968)

(Disabled sample) Hours worked on previous job	16
" " Duties required in prev. job	10

Parry and Crossley (1950)

Having registered and voted	23
In possession of a library card	10
Ownership of a driving licence	10
Home ownership	4
Age	17

It was decided to base estimates of the invalidity of the facts cited in the parental interviews on comparisons with particular variables from the above list:

Table 25. Validity of parent interview variables

Variable	Assumed equivalent to:	% invalid ( $1 - r_{ii}^2$ )	Validity correlation
Reading behaviours	Registration and voting	23	.85
Language environment	Having library card	10	.95
Parent reading attit.	Registration and voting	23	.85
Mathematics behavs.	Having library card	10	.95
Parent-child cooperatn.	Contrib. to Community Chest	40	.77
Television time	Having library card	10	.95
Television behaviours	Registration and voting	23	.85

The comparison behaviours chosen for the reading and mathematics behaviours

differed on the grounds that a parent is likely to be more aware of having some responsibility in the area of reading than in that of number skills and thus the responses in relation to reading would be likely to be more biased.

The relatively high validity attributed here to the variable of television viewing time (this variable being based on a judgement made by E in the course of each interview) is partly borne out by a small study undertaken during seven interviews - again the sample was selected on the basis of E's judgement that the parents would not object to making a frank assessment of the time their children spent viewing. The figures below present the comparisons between E's coded estimates, made in the course of the interviews, and the parental estimates made at the end of the interviews after the matter had been explained to them.

Table 26. Comparison of estimates of children's viewing time

Family	Period	E's coded estimates	Parental estimates
A	Week-day	2-3 hrs	Approx. 3 hrs
	Week-end	More than 3 hrs	Maximum of 3 hrs
B	Week-day	1-2 hrs	1-1½ hrs
	Week-end	0-1 hr	¾ hr
C	Week-day	1-2 hrs	1½ hrs
	Week-end	0-1 hr	Very little
D	Week-day	1-2 hrs	1-2 hrs
	Week-end	1-2 hrs	1-2 hrs
E	Week-day	2-3 hrs	2-3 hrs (2 hrs every evening, plus 1-2 hrs some mornings)
	Week-end	1-2 hrs	2 or maybe 3 hrs
F	Week-day	1-2 hrs	1 on average (parent's description of range of programmes watched suggested more)
	Week-end	2-3 hrs	1 hr on average (description as above)
G	Week-day	2-3 hrs	2-3 hrs
	Week-end	2-3 hrs	Over 3 hrs

The final reliability figures for each parental variable are determined from the product of the adjusted rater reliability figures with the validity correlation figures. All the figures are presented in table 27 (overleaf).

The crudeness of these estimates has to be recognised. At the same time the point needs to be reiterated that there is no credible alternative to recognising that both the rater reliability and the validity of parental reports reduce the accuracy of the variable scores. Making an assessment of these sources of

Table 27. Parent interview reliability estimates

Variable	Adjusted rater reliability	Factual validity estimate	Combined reliability estimate
Reading behaviours	.80	.85	.68
Language environment	.95	.95	.90
Parent reading attitude	.95	.85	.81
Mathematics behaviours	.90	.95	.85
Parent-child cooperation	.60	.77	.46
Television time	.90	.95	.85
Television behaviours	.60	.85	.51

inaccuracy, however limited the precision of the assessment, and taking statistical account of that inaccuracy in the disattenuation process, are legitimate strategies in a project whose conclusions are based largely on the interrelationships between a considerable number of variables of differing degrees of reliability.

#### 6.515 Reliabilities of other data

The reliabilities of 37 test and interview variables have been described in the preceding sections. The remaining variables consist of a selection of categorical and interval measures. In the case of the categorical measures, such as sex, ethnic group and the different programme groups, a reliability of 1.0 has been assumed, although the confusion between two parents of the same name but different ethnic group at one school indicated how even such certainties as categorical variables could have been wrongly recorded had the study been a brief one.

There were five age variables: one absolute 'sample age' recorded for the sample as a whole, on a fixed baseline date, and separate age scores for each of the four testing occasions (in the case of very young children the nursery attainment and nursery ability tests were given some months apart). A measuring device was used to assess these age spans to the nearest half month. It was considered that with the possibility, however remote, of either the school or E recording a slightly incorrect birth data, and the inevitable problems of accuracy in scoring nearly a thousand age totals in half months, a reasonable figure for all the age reliabilities would be 0.98.

The two variables 'length in nursery' and 'length in reception class' are discussed in section 6.60 in some detail. There was more uncertainty about the accuracy of these estimates than there was about the age estimates, since an attempt was made to correct the nursery and school period scores for the amount of child absence from the nursery or main school, making reference as much as possible to the school registers. The issue was complicated by the varying periods

during which some children attended full-time nursery (a few of the schools having introduced this change near the end of the children's nursery periods). It was decided to estimate these reliabilities at 0.96 each, although in retrospect it may have been more accurate to estimate the nursery figure at a slightly lower level and the reception figure at a higher level.

There were four variables on parent programme attendance. Two of the variables scored the actual attendance at reading or mathematics meetings, respectively; there were two other variables in which each meeting attendance was weighted by E's estimates of the degree of which environmental disturbance (such as children running through the meeting area frequently, in three of the venues, or other problems of infant 'intervention') had reduced the effectiveness of the meeting. For the variable which directly recorded programme attendances the reliabilities were estimated at 0.97. It was judged that the remaining two variables (weighted attendances) had a factual validity of about 80 per cent, giving a reliability figure of 0.89. (Realities such as late-coming reduced the accuracy of all four scores.)

It is worth noting at this point that the above reliability adjustments did not necessarily result in higher prediction levels (within the models) for the weighted attendance figures than for the unweighted figures. It is possible that E's subjective estimates of the effects of environmental disturbance were too high, or alternatively that the correction for factual validity was too severe. The point is emphasised to show that the disattenuation process does not automatically favour the more highly disattenuated variables in a multivariate relationship; the correction is simply an attempt at statistical honesty in so far as reliability can be assessed, without regard to whether disattenuation will increase or decrease the contribution of a particular variable in relation to other predictors. It is the same consideration which underlies the warnings of Cronbach et al (1972) on the need to take statistical account of unreliability in the predictors.

The remaining four interval measures were the two need for esteem and two need for security variables, assessed by nursery teachers in the nursery classes and by reception teachers in the reception classes. The only possible measure of reliability here was to ask nursery teachers and their two assistants in each of two large nursery classes to carry out separate assessments on other children. The assessments were made on a total of 30 children who had been attending for at least three months. This gave three separate assessments on each of the children.

The three reliability figures for need for security were .51, .48 and .15, giving a root mean square of 0.41. For need for esteem the comparable figures were .69, .50 and .68, giving a root mean square of 0.63.

These figures are extremely 'soft' since differing adults can have widely differing interpretations of levels of need and of the concepts themselves. The figure for need for security was so low as to raise some doubts as to whether the

variable should be used. It was allowed to remain, however, in view of the finding in the author's previous study that both these needs variables were related to reading levels nine months later.

#### 6.516 Reliabilities of latent variables

The method of path analysis used in the present study relies heavily on the creation of latent variables in different conceptual areas such as attainment, ability levels and home environment. In some cases these latent variables are used in regressions alongside raw variables which have been corrected for disattenuation.

The question arose of whether the latent variables should also be disattenuated prior to the regressions. Since this type of variable is constructed from a set of unreliable variables, it was considered necessary that each latent variable should have its own reliability parameter, based on a weighted sum of the reliabilities of its constituent variables, the weights being the regression coefficients used in the construction of the latent variable (LV).

This procedure was followed throughout, using the derivation

$$r_{LV.LV} = \frac{\sum_{i=1}^{i=m} (b_i r_{ii})}{\sum_{i=1}^{i=m} b_i} \quad i = 1, 2, \dots, m$$

where  $r_{LV.LV}$  is the reliability of the latent variable,  
 $b_i$  are the weights (regression coefficients) used to construct the latent variable,  
 $r_{ii}$  are the reliability coefficients for each variable used in forming the latent variable, and  
 $m$  is the number of constituent variables.

(The possibly more accurate alternative of deriving

$$\left[ \sum (b_i \cdot r_{ii}^2) / \sum b_i \right]^{1/2}$$

gave results which usually appeared to differ very little from those obtained when the original derivation was used with sample sets of coefficients and reliability values. In view of the uncertain implications of introducing square root derivations of a statistic based in part on regression coefficients, it was considered preferable to rely on the simple weighting method used in the first derivation above.)

## 6.52 Validity determination

This section will deal mainly with the nomological validity of the original variables used in the study, as well as with the nomological redundancy of the regression equations used for creating the latent variables. Redundancy statistics on the three principal equations used in each path model will be presented in section 6.80.

The content validity of the variables has already been defined in the description of the variables themselves. Other aspects of validity, such as the concurrent and construct validity, will not be dealt with specifically except in the case of the newly developed Infant Reading Test. While both these latter forms of validity offer useful insights into the nature of particular variables, it is considered that for the purposes of the present study the nomological validity - or the additive validity of individual variables and the redundancy index of equations built on those variables, in the context of a set of predictors in the same conceptual field - provides a better measure of credibility.

Section 5.15 dealt in detail with the theoretical rationale for nomological validity and set out the proposed statistical expressions of this concept. Briefly, it is argued that a variable's validity needs to be assessed within its context as one of a number of predictors of some outcome variable, rather than in isolation. A highly 'valid' relationship between a variable and some outcome will not necessarily continue when other more powerful variables are used in the same model. Thus it is the nomological validity, or validity within a model, which should be the criterion of judgement.

Clearly this concept can be abused if a large number of disparate variables are entered into a regression equation at the same time. Nomological meaningfulness at the most basic level rests on the idea of a set of conceptually similar variables being employed in an equation to predict an outcome. Thus the validity of a particular variable needs to be judged in relation to a set of comparable variables. At a higher level it is of course possible to examine the validity of more global variables (such as 'ability' or 'home environment'), whether these are simple additive entities, latent variables weighted according to regression coefficients, or other forms of clustered assemblies. Here a number of global variables can be entered into joint predictive equations and their additive validities compared in the same way as is done with the single variable predictors. At this higher level one will be comparing the additive validities of latent or grouped variables in relation to some outcome assumed or hypothesised to be based on all these groupings.

Section 6.80 discusses the nature of the initial conceptual groups of variables and of the ultimate models containing those groups when assembled in the form of latent variables. The present section examines the nomological findings on the individual and group variables; these are looked at in relation to the study's major criteria of reading and mathematical performance at the end of the field research phase.

Essentially the question is asked whether the variables used in the analysis play a meaningful part within their particular networks of relationships, contributing uniquely towards the total variance explained. The nomological criterion of additive validity compares the size of the variable's unique contribution to the outcome variance with the size of the squared correlation between the predictor variable and the outcome. The other major nomological criterion, known as the redundancy index, offers a combined statistic based on the lack of predictive power in the equation, the invalidity level of the constituent variables, and the predictive shortfall of the variables within that equation.

In a certain sense all the validity criteria are also a reflection of the strengths and weaknesses of the models developed in the study; the redundancy index in particular - and its three constituent parameters - offer specific evidence on the statistical frugality or otherwise of the assemblage of variables within a model.

The Appendices D7 (1 to 10) present the additive validity and other parameters for each of the original variables used in the development of the latent variables for the ten path models. Those tables present the single contributions ( $r^2$ ) between each predictor variable and the particular outcome variable, the unique variance contribution to that outcome, the additive validity and the regression coefficient for the predictor. The table presented in this section (overleaf) summarises the additive validity data from the Appendices.

This material will be reviewed below. It should be remembered that since additive validity is the simple measure of  $r^2$ , as defined above, divided by the unique variance contribution of the variable to that outcome, this statistic does not necessarily reflect the predictive power of a variable in relation to other variables, but rather the sensitivity or specificity of its additive contribution to the model. Thus a variable having a high correlation with the outcome but making only a small unique variance contribution to the regression equation is not a highly valid variable, in nomological terms, since most of its variance is either shared with other variables or excluded by other more 'valid' variables. On the other hand a variable having only a low correlation with the outcome, but yielding much or most of that correlation in the form of unique predictive variance, is a highly 'valid' contributor, despite the limited size of its contribution.



Table 28. Nomological validity comparisons for variables used in regressions

Group and variable names	No. of models in which vbl. appears* (out of possible 10)	Mean Additive Validity	Highest Add. Validity
<u>Nursery Ability</u>			
1. WPPSI Inform. Nursery	10	0.187	0.371
2. " Sentences Nurs.	9	0.200	0.383
3. " Pic.Compl. Nurs.	10	0.209	0.376
4. " Block Des. Nurs.	8	0.118	0.227
5. Rhythmic Tapping Nur.	8	0.222	0.644
6. Match. Fam. Figs. Nur.	8	0.230	0.393
7. Bender Gestalt Nurs.	7	0.136	0.329
8. Self-picture Nursery	0	-	-
9. Distractibility Nurs.(-ve)	3	0.152	0.213
10. Sex Group Nursery**	3	0.445	0.625
11. Age Nursery Assessment	8	0.360	0.515
<u>Initial Attainment</u>			
1. Reading awareness Nur.+	7	0.202	0.421
2. Infant Reading Test N.+	9	0.291	0.458
3. Maths Numeracy Nursery+	10	0.357	0.509
4. Maths Concepts Nurs.+	9	0.179	0.367
5. Piagetian Tests Nurs.+	9	0.110	0.372
<u>Parent Academic Environment</u>			
1. Reading behaviours	7	0.431	0.565
2. Language environment	1	0.103	0.103
3. Parent reading attitd.	0	-	-
4. Maths behaviours	8	0.338	0.493
5. Parent-child cooperatn.	0	-	0
6. TV viewing time (-ve)	1	0.500	0.500
7. TV controlling behavs.	1	0.638	0.638
<u>Parent Programmes</u>			
1. Reading meetings attended	2	0.255	0.289
2. Reading meetings weighted	1	0.198	0.198
3. Maths meetings attended	1	0.144	0.144
4. Maths meetings weighted	2	0.446	0.617
Legend and further variables overleaf			

Table 28 (continued)

Group and variable names	No. of models in which vbl. appears* (out of possible 10)	Mean Additive Validity	Highest Add. Validity
<u>Reception Ability</u>			
1. WPPSI Inform. Receptn.	9	0.137	0.258
2. " Sentences Receptn.	9	0.205	0.334
3. " Pic.Compl. Recep.	9	0.124	0.228
4. " Block Des. Recep.	10	0.130	0.309
5. Rhythmic Tapping Recep.	7	0.183	0.364
6. Match. Fam. Fig. Recep.	9	0.230	0.488
7. Bender Gestalt Receptn.	8	0.111	0.231
8. Self-picture Reception	1	0.047	0.047
9. Distractibility Recp.(-)	2	0.251	0.277
10. Sex Group Reception**	2	0.210	0.254
11. Age Reception Assessmt.	8	0.378	0.463
<u>Nursery Needs</u>			
1. Need security Nursery	4	0.262	0.432
2. Need esteem Nursery	8	0.465	0.810
<u>Reception Needs</u>			
1. Need security Reception	6	0.284	0.429
2. Need esteem Reception	9	0.539	0.736
<u>Single Variables***</u>			
1. English Pic.Vocab. Nur.	1	0.099	0.099
2. Time in Nursery	7	0.242	0.385
3. Time in Reception	7	0.104	0.237
4. Age Post-Test Assessmt.	7	0.058	0.084

**Legend:**

- \* Variable included in the 'parent' models if the regression coefficient's probability level is sufficiently low, or the variable's unique variance (within that model) is sufficiently high, according to the criteria described in section 6.80.
- \*\* Sex is entered twice, in both the Nursery and Reception Ability groupings, as its contribution varies in relation to each group.
- \*\*\* For reasons discussed in section 6.80, these single variables are not entered as part of any of the above conceptual groups; the variables are, however, entered at appropriate points in the path models. The additive validity statistics are drawn from the variable's earliest appearance in the path models.
- + In one satellite model the regression equation is over-predicted; the validity figures are reduced accordingly. The adjusted figures are not used in any further derivations. The 'satellite' principle is discussed in sec.6.80.

6.521 Nomological validities of individual variables

The table on the previous two pages summarises the information on the nomological validities of each of the original variables assessed for this study; in each case these parameters are determined in relation to the variable's prediction of post-test attainment (reading, mathematics or both in combination). The information on the variables is repeated in a different form in section 6.80, where the specific contribution of these variables to each of the latent variables used for the ten path models is displayed in diagrams.

Nursery Ability variables

The four WPPSI scales perform well and make a meaningful contribution to most of the models at the Nursery level. For the first three scales (Information, Sentences and Picture Completion), the additive validity is close to one fifth, implying that in these ten models approximately one-fifth of the squared correlation with the dependent variable reappears as unique variance. One of the WPPSI variables performs rather poorly; Block Design at Nursery level makes no contribution to two models and its mean additive validity of 0.12 indicates a limited contribution overall. Surprisingly it fails to contribute within either the Advantaged Sample or the Disadvantaged Boys sample; in contrast Block Design makes an important contribution to the attainment scores of Disadvantaged Girls and Disadvantaged Black Children.

There are four further cognitive tests included under Nursery Ability. Rhythmic Tapping makes a relatively small but reasonably valid contribution to all but two models; these are the Advantaged Sample and the Disadvantaged Maths Programme Groups (in the latter case with maths concepts as the attainment outcome). However for the two samples specifically focused on reading attainment and maths numeracy attainment (Disadvantaged Reading Programme Groups and Disadvantaged Maths Programme Groups), Rhythmic Tapping has high additive validities of 0.42 and 0.64, with corresponding unique variance contributions of 3.5 and 12.3 per cent. This is an interesting finding.

(By way of shortening the nomenclature, the Disadvantaged Reading Programme Groups and Disadvantaged Maths Programme Groups will be described as Disadvantaged Reading and Disadvantaged Maths in the text. The importance of the three models based on Programme Groups is that these focus on the parents at the disadvantaged schools who took part in the early reading and early maths programmes; the other seven models include both Programme parents and Working Group parents, as well as the small minority of non-working parents who did not attend any programme meetings (22 per cent).)

In contrast to the results with Rhythmic Tapping, Matching Familiar Figures, while producing a comparable level of mean additive validity, fails to yield

predictive variance for the same two models where Rhythmic Tapping performs so well; for the Disadvantaged Reading model the M.F.F. has a relatively low correlation of 0.23 and an almost zero variance contribution to the reading outcome, and for Disadvantaged Maths (with the maths numeracy outcome), the correlation is -0.11, with a variance contribution of 0.4 per cent.

The Bender Gestalt has a lower mean additive validity, and fails to yield unique variance in the same two samples as does the M.F.F. - Disadvantaged Reading and Disadvantaged Maths (numeracy) - and also in a third sample, the Advantaged group. The similarities between the Bender and the M.F.F., in that both are thought to assess spatial abilities in the child, are evidenced by the specificity of the predictive validities of each test. At the same time they each make their own independent contributions to variance within the same equations, suggesting definite differences in the types of abilities tapped.

The Goodenough-Harris Draw-a-person test, here termed the Self-Picture Test, performs extremely poorly, failing to contribute sufficient unique variance to be included in any of the Nursery models. This is not related to a particularly low correlation; the relationship with the outcome variables ranges from 0.18 to 0.41 in the different samples. What this finding does imply is that, when combined with other indicators of cognitive functioning at Nursery level, the Draw-a-person has no independent or nomological validity in relation to reading or maths attainment.

It had been hoped that the Distractibility variable (scored negatively) would yield some useful information; the different attempts to quantify this measure have been described earlier in the study (section 4.31). Of the only three samples in which it made a useful contribution, two (Disadvantaged Reading and Disadvantaged Maths (numeracy) ) have already been noted for their differences on two of the other cognitive variables. The surprising inclusion here is the Advantaged Sample, where distractibility contributes 2.4 per cent unique variance (correlation 0.38).

The highest correlations in samples where Distractibility has been excluded are 0.26 for Disadvantaged Girls and 0.28 for Disadvantaged Black Children. An examination of the means for Distractibility shows that whereas the Disadvantaged White and Black Children have almost the same levels on this indicator, Disadvantaged Boys have a mean level a quarter standard deviation lower (i.e. more distractible) than that of girls, as well as a much higher variance in the level of distractibility. Despite this fact, the correlation of boys' distractibility with that of the total attainment outcome measure is only 0.05.

An important point that should be noted here and elsewhere in this analysis is that failure to contribute to variance does not necessarily point to either a generally low or generally high score of the variable concerned. What

it does indicate is the possibility:

- a. that the amount of variance shown by a variable is too small to contribute meaningfully to some outcome; or
- b. that whatever variance there is bears little relationship to the variance of the outcome variable; or
- c. that whatever variance is shared between the predictor and the outcome measure is overshadowed by a more powerful predictor.

Correlational evidence and data on means and standard deviations can help to suggest which of these alternatives is more credible. Although careful dissection of the inverse correlation matrix - on which the regression parameters are based - might offer yet more conclusive evidence, this is a tricky procedure, given the known fallibility of the correlations themselves prior to inversion.

The Sex Group variable is another of the variables in this set which makes little contribution to overall attainment; it does not appear in any of the samples with the combined outcome. However in the three samples focused on particular outcomes - Disadvantaged Reading (with post-test reading as the outcome) and Disadvantaged Maths (numeracy and concepts are two separate outcomes) - the Sex Group variable does make useful contributions and shows high additive validity. The correlations of Sex itself with the six individual outcome variables (the three reading tests, maths numeracy, maths concepts and the Piagetian tests) do not show any strong sex bias; in general Sex (scored 0 for boys and 1 for girls) yields a very small positive correlation (from 0.12 down to 0.01) for all outcome variables except maths numeracy, where there is a slight negative correlation of  $-0.03$ , indicating that if in nothing else, boys did at least outshine girls in numeracy.

Age of Nursery Assessment is an important variable and appears in eight of the models with a particularly high additive validity of 0.36. It is potentially an interesting predictor, since this is not a measure of the correlation of post-test age with post-test attainment, but rather a measure of the age at which this battery of tests was administered to the individual children in the Nursery and its relationship to the post-test outcome. By its very nature it serves as an age 'monitor' for the test scores, so that its contribution to the final outcome takes account of the varying ages of the children when they performed these tests. It was seen as a more sensitive and useful indicator of the age variable than it would have been if all the scores had been simply standardised for age. Its unique variance contribution ranges from 1.3 to 8.3 per cent. The only two samples where it fails to appear are in the Advantaged Sample and the Disadvantaged Maths (numeracy) groups.

While it is possible to argue that the absence of Age from numeracy predic-

tion suggests that very few sample children have any awareness of numbers at Nursery level, the absence of Age from the prediction of total attainment for the Advantaged Sample needs particular attention. In the Advantaged Sample there were a handful of very bright children who in general were younger than the class average. It is likely that this fact is responsible for the failure of the Age variable to predict outcome variance; corroboration is found in the negative correlation of  $-0.11$  between Nursery age and total attainment for the Advantaged Sample, followed eighteen to twenty-two months later by a fairly similar correlation of  $-0.13$  between Post-Test Age and post-test attainment for the same sample.

While this reverse correlation is part of the reality of an action research situation, it does serve as a warning of the peculiarities of small samples (the Advantaged Sample having only 27 children). In a larger sample such biases are less likely to occur. At the same time it should be pointed out that the unique variance contributed by the Nursery Age variable (within the Advantaged Sample model) is only 0.1 per cent, so that the negative age relationship in no way affects the model as a whole.

#### Initial Attainment variables

It would be surprising if the Initial Attainment variables did not feature prominently in the prediction of post-test attainment, since the Nursery battery contains four out of the six tests used again for the post-testing. As can be expected, the Infant Reading Test variable does not feature in the prediction of post-test maths numeracy (for Disadvantaged Maths), and likewise Maths Concepts at Nursery level does not predict post-test reading. On the other hand the post-test maths concepts outcome is predicted jointly by the two Nursery maths tests (Maths Numeracy and Maths Concepts), by the Piagetian Tests and by the Infant Reading Test, although the latter contributes only 1.2 per cent out of a total predicted variance of 50.9 per cent. The nomological parameters show that Maths Numeracy (Nursery level) has the highest mean additive validity in this set; its unique contributions to total attainment are approximately the same as those of the Infant Reading Test. The Reading Awareness variable fails to appear in either of the two maths models, again in keeping with expectation. However it also fails to appear in the Advantaged Sample's predictive set.

Although the mean score for Reading Awareness for this sample is 6.5 (s.d.2.0) compared to a Total Disadvantaged Sample mean of 5.6 (s.d. 2.4), the low correlation of the Advantaged Sample's Reading Awareness with post-test attainment (0.21, compared to 0.39 for the Disadvantaged Sample) is at the root of this low validity. The low predictive power may also be related to the fact that most children in the Advantaged Sample appeared to have some awareness of the purpose and use of books by the time they reached the Nursery class, whereas

any variation in the parents' academically oriented behaviours in the disadvantaged samples' homes were so low as to yield minimal predictive power for post-test attainment. The data bear out this impression. Parents' Reading Behaviours in the Advantaged Sample have a mean level of 12.6 (s.d. 3.7), compared to a mean of 7.8 (s.d. 3.5) for the Total Disadvantaged Sample; for Mathematics Behaviours the mean is 11.5 (s.d. 3.8) for Advantaged parents, compared with 3.4 (s.d. 2.0) for Disadvantaged parents.

One other unusual fact in this variable set is that Mathematics Behaviours correlates only 0.20 with the maths numeracy outcome (in the Disadvantaged Maths groups), and fails to contribute uniquely to that outcome; in contrast this same variable correlates 0.34 with post-test maths concepts (for the same sample) and makes a useful 5.2 per cent contribution to that outcome. The implications of this finding will be discussed when the path models are reviewed.

#### Parent Programme variables

In examining this set of variables it should be remembered that most of the samples identified for the analysis consist of a mixture of programme attenders and those parents who were either working or who had failed to attend any programme meetings. Furthermore, no parent attended both reading and mathematics programmes, thus further reducing the possible prediction of the Programme attendance variables within the wider samples. Consequently, in each of the general samples between two-thirds and three-quarters of the sample have not attended one or other of the particular programmes.

Only four of the ten samples can therefore justifiably be examined in relation to the Programme variables, namely the Advantaged Sample (a very high proportion of this sample attended either reading or maths programmes), the Disadvantaged Reading Programme Groups and the Disadvantaged Maths Programme Groups (the latter sample featuring separately in two models predicting maths numeracy and maths concepts respectively).

For the Advantaged Sample, predicting total attainment, the Reading Meetings Attended and Reading Meetings Weighted variables each contributed moderately to the outcome measure. (Reading Meetings Weighted, as explained earlier, is a weighted score of meetings attended, based on the interruptions and external problems faced in each meeting - such as the presence of very restless or crying toddlers brought by the parents). The two remaining variables, Maths Meetings Attended and Maths Meetings Weighted, though each have a small correlation of 0.18 with this total attainment outcome, do not contribute uniquely to its variance.

For the Disadvantaged Reading groups, the prediction of post-test reading

a considerable minority in the Total Disadvantaged Sample had only the vaguest concept of what books were about when they arrived at the Nursery.

#### Parent Academic Environment variables

With the exception of Reading Behaviours and Mathematics Behaviours, the variables in this group make little or no contribution to final outcome. The Language Environment variable appears only as a predictor of post-test reading (for Disadvantaged Reading groups); TV Viewing Time (scored negatively) makes a handsome contribution to the maths concepts model, while TV Controlling Behaviours contributes moderately to the Advantaged Sample's outcome. These are however single contributions; in nine other models these three variables fail to provide any useful variance. The Parent Reading Attitude and Parent-Child Cooperation variables do not feature in any of the models; the former variable has little or no correlation with post-test outcomes, except in the case of the Advantaged Sample where its correlation of 0.35 fails to yield unique variance beyond the level of 0.1 per cent.

The performance of this variable set merits some attention, in view of the importance attached to the parent environment in the development of children's early academic attainment.

Parents' Reading Behaviours appears in seven of the models but fails to appear as a predictor in the models focused on maths numeracy and maths outcomes; this is to be expected. However, it also fails to appear in the Disadvantaged Black Children's model, where the post-test outcome is total attainment. Maths Behaviours also fails to appear as a predictor in this latter sample. For the Disadvantaged Black Children's sample the correlations of the parents' Reading and Maths Behaviours are so low - approximately 0.13 for each variable - and the colinearity of Reading Behaviours with another low predicting variable, Language Environment, is relatively so much higher (at 0.43) that the Parent Environment equation yields a negative prediction. This is a statistical quirk which can arise where adjustments are made to correct  $R^2$  for the numbers of cases and variables in a model; however it is an oddity which is only likely to occur when the overall unadjusted prediction is of the order of a few per cent. (The rationale for the adjustment is presented in section 5.20.)

While most of the samples have a rather limited level of prediction for the set of Parent variables, it is noteworthy that in the Advantaged Sample these variables yield the high prediction of 26.5 per cent of post-test variance. While it could be argued that the quality of the assessment of advantaged homes in the parent interviews may have been superior to the assessment in the disadvantaged homes, a more disturbing and more likely possibility is that the level of



attainment yields only a limited unique variance when both Reading Meetings variables are entered jointly; as the weighted reading attendance variable shows zero prediction in the presence of Reading Meetings Attended (the latter correlating 0.24 with the outcome), it is this latter variable which is used as the Parent Programmes latent variable in the further analysis of the model.

The Disadvantaged Maths sample yields modest contributions for the Maths Meetings variables, in predicting both maths numeracy and maths concepts. In the latter case only Maths Meetings Weighted predicts uniquely, and this has been retained for further analysis of the model. In the former case both variables have been used to construct the Parent Programmes latent variable.

#### Reception Ability variables

The additive validity characteristics of these variables bear a certain similarity to those of the same variables measured at Nursery level. It is informative to note the ways in which some of the variable parameters differ, however.

The WPPSI Picture Completion has a smaller validity figure at Reception level, possibly because some of the children were reaching close to ceiling on a set of scales where the maximum test age is  $6\frac{1}{2}$  years.

Rhythmic Tapping remains an interesting variable. It is still absent from the prediction of maths concepts and likewise from the Advantaged Sample and the Disadvantaged Working Group (both these samples predicting to total attainment). The minimal unique variance contribution of Rhythmic Tapping in the latter sample at Nursery level has disappeared entirely at Reception level. As discussed elsewhere, it was found that the Disadvantaged Working Group appeared to be upward mobile, with their children on the whole performing better than the children of mothers who did not go out to work. (See section 6.431 and a further discussion of the issue in section 6.80.) While the parallels between the Advantaged Sample and the Disadvantaged Working Group are not too close, there are similarities such as that described above which should be noted.

A particular feature of Rhythmic Tapping as a predictor is that its importance at Reception level has increased considerably in the case of Disadvantaged Girls and Disadvantaged Black Children, while it shows a particularly high prediction of 10.1 per cent in the case of the Disadvantaged Reading groups' model for predicting reading attainment. With some of the other samples the predictive strength has declined. This could of course be considered a statistical artefact, or an example of the low reliability of the measure. The reliability is certainly limited (test-retest figures gave a correlation of 0.58 with a young sample); on the other hand the 9-month correlation between rhythmic tapping at the Nursery and Reception levels, over the whole sample, is 0.65. (This figure compares well with the nine-month correlations between the Nursery

and Reception levels of the WPPSI scales, which range from 0.77 for Information to 0.56 for Block Design.

The Self-Picture (Draw-a-person) test again makes almost no contribution to outcome variance; it appears only once, in the Disadvantaged Maths groups, making a contribution of 0.9 per cent to numeracy. Reception level Distractibility and the Sex Group variable (re-inserted into this set of predictors as a concurrent influence) both make low and mostly negligible contributions, as occurred previously at Nursery level. The only differences are that both these variables now make modest contributions to the maths numeracy criterion.

Age of Reception Assessment continues to make its important and highly valid contribution, again with the same exceptions as appeared at Nursery level. It should be noted that this variable measured at Reception level has a correlation of only 0.79 (62 per cent shared variance) with Age of Nursery Assessment. The absence of a close relationship has arisen because it was not possible to undertake all the Reception class assessments in the same order and over exactly the same intervals for every child in each of the six classes, for technical reasons concerned with differing school holidays and similar factors.

#### Nursery Needs and Reception Needs variables

These pairs of variables, based on teachers' assessments in the Nursery and Reception classes, are entered separately from the other variables obtained from the children, since the Needs assessments are considered to have a very different conceptual meaning from either Ability or Attainment.

Despite this precaution the Needs assessments make relatively limited contributions to the variance in most samples. However the situations in which they do make a contribution appear to be of some importance.

In nearly all the samples it is Need for Esteem which makes the main contribution to outcome variance, rather than Need for Security. This occurs at both the Nursery and Reception levels. The Disadvantaged Boys sample shows a fairly strong prediction for Need for Security at Nursery level, this contributing 3.5 per cent to outcome variance. Need for Esteem at Nursery level is a consistently good predictor, rising to a maximum contribution of 11.9 per cent of outcome variance in the Disadvantaged Black Children's sample; while this may be an example of a 'good' predictor, there are disturbing implications for the finding and these will be examined in some detail in section 6.80.

One sample which yields no prediction whatever for Nursery level needs is the Advantaged Sample. High colinearity problems and a low predictive variance also prevent the two Nursery Needs variables from making a joint contribution to the outcome for Disadvantaged Reading. The main predictor, Need for Esteem at Nursery level, has been retained in the path model for this sample

because of its correlation of 0.27 with post-test reading.

Need for Security at Reception level makes a stronger contribution than at Nursery level, both in terms of the models where it has been retained and in its mean additive validity. It is again absent from the Advantaged Sample, and also, somewhat surprisingly, from the Disadvantaged Boys sample, since this variable was a strong predictor at Nursery level; it is also missing from the Disadvantaged White Children's sample.

The last of these variables, Need for Esteem at Reception level, appears as a particularly strong variable, contributing as much as 12.7 per cent of the outcome variance in the Disadvantaged Black Children's sample. Both the Reception Needs variables fail to appear in the Disadvantaged Maths prediction of numeracy, due to high colinearity and a very low prediction.

Although the additive validity for these four variables appears comfortably high, it should be realised that since each variable has only been in competition with one other variable, it is inevitable that its nomological validity should appear to be quite high. The Redundancy Index statistics to be presented in the next sub-section will make the shortcomings of these limited models more apparent.

### Single Variables

The legend under Table 28 explains why these variables are assessed as part of their respective path model groupings and not as part of any of the initial conceptual groupings.

The nomological validity of the variables is assessed in the context of their initial contributions to the path models themselves, rather than to the earlier regressions from which the latent variables were formed. The reasons for this need a brief explanation.

The variables Time in Nursery and Time in Reception are the measures of the time spent by each child between the date of the Nursery assessment and the date of leaving the Nursery class, and of the full period spent in the Reception class between the date of entering the class and the date of the post-test assessment for that child. As described in section 6.60, both variables have been 'school-centred' so that they yield time measures linked to the mean period of nursery and reception attendances in their own classes; this eliminates variable biases caused by the testing rota (schools being visited in the same order during each testing phase).

Clearly the time variables cannot fit into any conceptual framework other than that of time itself - at the two different levels at which these variables are measured.

Age of Post-Test Assessment is likewise a single variable which does not appear to fit into any conceptual category, since it is measured at the same time as the post-test assessment itself; its purpose is to extract any further variance due to age which has not already been accounted for by the two previous measures of age - namely Age of Assessment at Nursery level, and again the parallel age at Reception level.

The variable English Picture Vocabulary Test is a most unusual variable in the present context. Despite its positive correlation with all the outcome measures in the various Disadvantaged samples, ranging from 0.27 with maths numeracy to 0.52 with total attainment for the Working Group (and a high correlation of 0.76 with total attainment for the Advantaged Sample), the variable contributes either negatively or not at all when included with the Nursery level Initial Attainment measures in the prediction of any of the disadvantaged samples' outcomes. In an effort to crystallise its contribution and free it from any preliminary conceptual grouping, and in view of its unique importance as a recognised measure of language ability, it was decided to enter the variable as a separate contributor to the main path analyses.

Surprisingly, even in this form it still makes virtually no contribution to the outcome variance in any of the nine disadvantaged samples. Only in the case of the Advantaged Sample does it make an abiding contribution to outcome variance at each of the three stages in the path model, ranging from 5.7 per cent of the outcome variance at Nursery level to 2.5 per cent at the full post-test level (where it is in competition with all the other major predictors). Even at the Nursery level, despite its correlation of 0.76 with the outcome measure, this variable has only a modest additive validity of 0.99.

The absence of prediction in the Disadvantaged samples is a most puzzling finding. The high correlations with the outcome measures are evidence that the variable was accurately measured; inaccurate marking of the children's scores would simply have reduced both the correlations and the regression parameters. The fact that the variable does contribute well in the Advantaged Sample, as compared with zero or negligible predictions in all the Disadvantaged samples, suggests that in the latter samples the E.P.V.T. measure is largely duplicated by other more powerful variables, in particular Initial Attainment and Ability. Although reasonably present in correlational terms (so that it would offer a prediction for the Disadvantaged outcomes in the absence of any other variables), this vocabulary assessment is clearly not strong enough to make its contribution within a complex model in these samples.

The finding also suggests that vocabulary skills in disadvantaged schools may play a much less prominent part in academic development in these early years than they do for advantaged children. This may be linked indirectly to the

finding of Tizard, B. et al (1980) that although the home language experience of working class and middle class samples of nursery age children did not vary noticeably across social class, there was a noticeable class difference in the richness of the experience faced by these children in the nursery class setting. Whether this is due to the cultural distance and the differences in language usage separating relatively advantaged teachers from their disadvantaged pupils, or whether it is due to shortcomings in the quality of language skills brought to school by many of the disadvantaged children, is a basic issue which could not be tested from the present data.

By way of a concluding comparison, the mean E.P.V.T. score for the Total Disadvantaged Sample is 10.8 (s.d. 15.4), compared to a mean of 14.9 (s.d. 18.2) for the Advantaged sample.

The contribution that E.P.V.T. scores make to the outcome variance in the Advantaged Sample will be discussed when that sample's path model is examined.

The three remaining single variables, namely Time in Nursery, Time in Reception, and Age of Post-Test Assessment, each contribute to seven of the path models. In a most interesting reversal, it is found that none of the three variables make any contribution to the Advantaged Sample's outcome of post-test total attainment. The explanation for the slightly negative relationship with age in this sample has already been given. It is not easy to translate this same explanation into a possible answer for the negligible relationships with Time in Nursery and Time in Reception (correlations of -0.03 and -0.11 with the outcome), since one would expect that whatever the age and brightness of the children there should be some tangible relationship between the times spent in the nursery and reception classes and the ultimate performance in reading and mathematics.

What may be occurring is that the contribution of the parents' home environment, in academic terms, is so powerful in the early years that it totally outweighs the contribution of the school - in many cases the latter may be offering a more sophisticated replication of what the child already knows as a result of its academically-oriented upbringing. This would not be an unusual finding for an advantaged school. It has been claimed on occasion that for the child from an advantaged and academically aware home, the first year at formal school is largely occupied with learning to mix socially and to adapt to the group situation in which the child is no longer one of the few or only centres of adult attention.

Two other situations in which these three 'time' variables make no real contributions (despite modest correlations varying from 0.24 to 0.34) occur in the two models based on Disadvantaged Maths groups, predicting numeracy and maths concepts. Here it may be argued that at nursery level and in the first

school year there are only the early beginnings on the teaching of maths concepts and numeracy, so that despite the modest and positive correlational evidence that Time in Nursery and Reception and Age at Post-Testing all count, they do not count sufficiently to make an independent contribution to variance.

Many of the issues arising from these pages will be referred to in context when the path models are examined in section 6.80.

#### 6.522 Nomological validity parameters for regression models

Table 29 on the three succeeding pages presents the redundancy index and associated parameters for each of the latent variables created for the ten path models developed in this study. Details of the sample, the sample size and the dependent or outcome measure in each model are set out in this table. For each model the same seven latent variables have been formed from the sets of conceptually similar variables described in table 28.

As the statistics presented in the following table are to some extent a forerunner of the path models developed and discussed in section 6.80 - and illustrated by appropriate path diagrams - the contents of this table will not be reviewed further at this stage, other than to comment on some broad features of the nomological parameters set out here.

Since the latent variables have been constructed to eliminate poor performing contributors, the nomological characteristics in general appear to be fairly satisfactory. The Variable Invalidity parameters vary somewhat, as can be expected from regressions covering between two and eight predictors. The higher the number of predictors the lower, in general, is the additive validity of each contributing variable likely to be; correspondingly this implies a higher variable invalidity parameter for the model incorporating that set of variables.

The parameter of Unrealised Predictive Power (the summation of the squared correlations with the outcome, divided by the predicted variance) shows a reasonable range of performance, with only a few rather high figures.

It is the third parameter, the Predictive Shortfall of each regression equation, which shows the penalties of relying on minimally predicting variables. As mentioned in the previous sub-section, it is the low predictions of some of the Needs variable sets and a few of the other variable sets - in particular Parent Academic Environment and Parent Programmes - which produce exceptionally high predictive shortfall parameters, this in turn contributing to a high redundancy index for the regression equation. It should be noted that not all the Needs or Parent variable equations yield high shortfalls.

The redundancy index statistics show that in general most of the models

(continued on third page of Table 29)

Table 29. Nomological validity parameters for regression models

Model				Redundancy Parameters			
No.	Sample	Predictor Group	Dependent variable	Variable Invalidity	Unrealised predictive power	Predictive shortfall	Redundancy Index
1A	Total Disadvantaged Sample N=129	Nurs. Ability	Total Attainment (Post-Test)	5.02	2.79	1.99	1.248
1B		Init. Attainment		4.77	2.36	1.82	1.408
1C		Par. Acad. Env.		2.02	1.54	11.32	2.865
1D		Par. Programs		-	-	-	- +
1E		Recep. Ability		5.40	3.01	1.77	1.278
1F		Nurs. Needs		-	-	-	- ++
1G		Recep. Needs		2.07	1.38	10.13	2.675
2A	Advantaged Sample N=27	Nurs. Ability	Total Attainment (Post-Test)	3.48	2.09	2.08	1.326
2B		Init. Attainment		4.24	2.27	1.99	1.558
2C		Par. Acad. Env.		1.99	1.73	3.78	1.464
2D		Par. Programs		4.78	4.75	19.04	5.380
2E		Recep. Ability		5.04	2.96	1.68	1.431
2F		Nurs. Needs		-	-	-	- ++
2G		Recep. Needs		-	-	-	- ++
3A	Disadvantaged Working Group* N=28	Nurs. Ability	Total Attainment (Post-Test)	5.89	3.52	2.22	1.755
3B		Init. Attainment		-	-	-	- +++
3C		Par. Acad. Env.		2.66	2.03	7.81	3.047
3D		Par. Programs		No programme meetings attended			-
3E		Recep. Ability		6.71	3.48	1.74	1.621
3F		Nurs. Needs		-	-	-	- ++
3G		Recep. Needs		4.15	3.25	20.14	4.911
4A	Disadvantaged Reading Program Groups N=48	Nurs. Ability	Reading Attainment (Post-Test)	3.96	2.55	3.82	1.573
4B		Init. Attainment		3.08	1.73	2.36	1.434
4C		Par. Acad. Env.		7.36	4.80	30.62	5.885
4D		Par. Programs		-	-	-	- ++
4E		Recep. Ability		5.30	2.75	2.18	1.378
4F		Nurs. Needs		-	-	-	- ++
4G		Recep. Needs		2.71	1.79	10.07	3.200

- \* Satellite models. Base model is Total Disadvantaged Sample (Model 1)
- + Predictive power too low to justify any of this group of variables being entered in the relevant path model
- ++ Only one variable in this group has sufficient predictive power; the relevant regression equation is discarded and the single variable is used on its own in the path model
- +++ Two poor predictors retained in this satellite model caused over-prediction and an inability to determine the redundancy parameters.

Table 29 (continued)

Model				Redundancy Parameters			
No.	Sample	Predictor Group	Dependent variable	Variable Invalidity	Unrealised predictive power	Predictive shortfall	Redundancy Index
5A	Disadvantaged Maths Program Groups N=32	Nurs. Ability	Maths Numeracy Attainment (Post-Test)	3.18	2.96	3.10	1.429
5B		Init.Attainmt		3.49	2.12	2.98	1.996
5C		Par.Acad.Env.		-	-	-	- ++
5D		Par.Programs		-	-	-	- ++
5E		Recep.Ability		7.41	3.63	2.49	2.258
5F		Nurs. Needs		-	-	-	- ++
5G		Recep.Needs		-	-	-	- ++
6A	Disadvantaged Maths Program Groups N=32	Nurs. Ability	Maths Concepts Attainment (Post-Test)	5.31	3.98	2.81	2.004
6B		Init.Attainmt		3.35	2.19	1.96	1.280
6C		Par.Acad.Env.		2.01	1.53	6.88	2.358
6D		Par.Programs		-	-	-	- ++
6E		Recep.Ability		5.21	3.16	1.81	1.321
6F		Nurs.Needs		-	-	-	- ++
6G		Recep. Needs		2.09	1.82	9.40	2.884
7A	Disadvantaged Girls* N=71	Nurs. Ability	Total Attainment (Post-Test)	3.78	2.56	2.01	0.894
7B		Init.Attainmt		5.19	2.39	1.91	1.554
7C		Par.Acad.Env.		4.67	3.57	59.48	6.208
7D		Par.Programs		-	-	-	- +
7E		Recep.Ability		5.80	3.27	1.90	1.504
7F		Nurs. Needs		-	-	-	- ++
7G		Recep. Needs		2.01	1.52	7.50	2.437
8A	Disadvantaged Boys* N=58	Nurs. Ability	Total Attainment (Post-Test)	6.14	3.16	1.93	1.549
8B		Init.Attainmt		4.59	2.38	1.76	1.348
8C		Par.Acad.Env.		2.01	1.53	6.02	2.228
8D		Par.Programs		-	-	-	- +
8E		Recep.Ability		4.95	2.96	1.79	1.184
8F		Nurs. Needs		2.42	2.04	13.07	3.474
8G		Recep. Needs		4.12	2.19	29.48	4.895

- \* Satellite models. Base model is Total Disadvantaged Sample (Model 1)
- + Predictive power too low to justify any of this group of variables being entered in the relevant path model
- ++ Only one variable in this group has sufficient predictive power; the relevant regression equation is discarded and the single variable is used on its own in the path model



Table 29 (continued)

Model				Redundancy Parameters			
No.	Sample	Predictor Group	Dependent variable	Variable Invalidity	Unrealised predictive power	Predictive shortfall	Redundancy Index
9A	Disadvantaged Black Children* N=44	Nurs.Ability	Total Attainment (Post-Test)	5.58	3.17	2.45	1.687
9B		Init.Attainmt		4.83	2.44	2.43	1.740
9C		Par.Acad.Env.		-	-	-	- +++
9D		Par.Programs		-	-	-	- +
9E		Recep.Ability		5.33	3.21	1.86	1.384
9F		Nurs. Needs		2.25	1.74	6.81	2.591
9G		Recep.Needs		1.98	1.52	5.14	2.044
10A	Disadvantaged White Children* N=62	Nurs. Ability	Total Attainment (Post-Test)	5.45	3.06	1.78	1.313
10B		Init.Attainmt		5.11	2.30	1.48	1.248
10C		Par.Acad.Env.		2.24	1.70	8.24	2.753
10D		Par.Programs		-	-	-	- +
10E		Recep.Ability		5.75	3.28	1.87	1.481
10F		Nurs. Needs		3.90	3.34	40.21	5.566
10G		Recep. Needs		4.04	2.19	31.88	4.956
* Satellite models. Base model is Total Disadvantaged Sample (Model 1)							
+ Predictive power too low to justify any of this group of variables being entered in the relevant path model							
+++ Predictive power too low to justify a regression equation. The relevant variables are however retained in the Base Model, where they have a modest predictive power. See text for further discussion of this point.							

perform reasonably well. Indices of over 5.0 are on the high side; it is worth noting that in such cases it is usually all three contributory parameters which show up the unsatisfactory nature of the model. On the other hand it should be emphasised that the retention of these equations within the overall path model has been decided upon in view of the importance of the information they contain; in some cases the latent variables concerned have continued to survive even within the more stringent demands of the competition in the path model, though in most cases latent variables derived from unsatisfactory equations have disappeared at an early stage in the development of the path model.

Again it needs to be emphasised that the redundancy index is a useful summary of how well an equation performs; it should never be an all or nothing rule for or against inclusion.

6.523 Specific validity assessments

For a variety of reasons a small number of variables were subjected to specific validity assessments.

Distractibility. Nursery teachers in two schools were asked to estimate the level of distractibility in 54 sample children (see question form, Appendix D10). Their assessments were correlated with the assessment made by E during the course of the Nursery tests. The correlation was found to be close to zero.

This raised the issue of whether the variable should still be used in the analysis. The fact that E's assessment of distractibility showed a test-retest reliability of 0.61 suggested that whatever characteristic was being judged, it was being estimated reasonably consistently, given the subjective nature of the judgement and the fact that a child's test behaviour patterns could vary considerably according to mood and the interaction with E on a particular day. It was also hypothesised that the high degree of freedom of choice enjoyed by nursery children meant that they were seldom in the position of being expected to sit on one seat for up to half an hour, as occurred when they were playing the variety of test 'games' with E. Thus nursery teachers may not have been in a position to assess the context-dependent kind of 'test distractibility' which E hypothesised would also reflect a tendency to distractible behaviour in classroom learning situations in the formal school.

Rhythmic Tapping. Here the nursery teachers in several classes were asked to estimate the rhythmic skills of 70 sample children (see same question form, Appendix D10). In this case the teachers' estimates correlated 0.35 with E's measurement of rhythmic tapping ability in the test of that name. This was considered a reasonable relationship, given that rhythmic skills (as assessed by the teachers when watching children during musical movement sessions) may be hypothesised to be only distantly related to the auditory motor integration skills which is thought to be assessed in the tapping test.

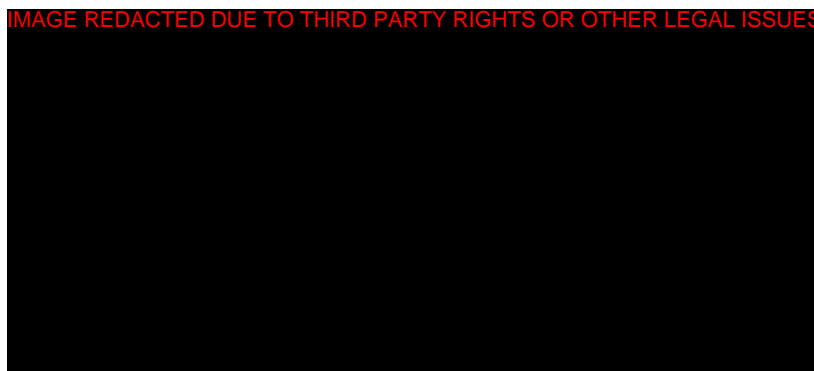
English Picture Vocabulary Test. Brimer and Dunn (1962) present a variety of data on the concurrent validity of the E.P.V.T. as a language measure. Correlations are reported of 0.73 with a measure of expressed vocabulary and of 0.76 with the WISC Vocabulary measure.

Wechsler Primary and Pre-School Intelligence Scale. Sattler (1974) reports that the WPPSI manual offers minimal evidence on validity. His own study cites a variety of evidence, including findings of congruent validities between WPPSI and Stanford-Binet measures which range between 0.33 and 0.92.

WPPSI Picture Completion Sub-Test. The results of this test showed the

expected distribution of scores in both the Nursery and Reception level assessments. There was however one item which caused a considerable problem for E in the marking, and later prompted a special study of the results on this item. Item number 15 (see below) shows what one assumes to be a frontal view of a car, pictured from road level perhaps five yards in front of the car.

15



(Copyright 1967, The Psychological  
Corporation, N.Y.)

Figure 16 . Item from the WPPSI Picture Completion Sub-Test

Although simply drawn, most of the usual fittings are present - the radiator, parking lights, windscreen wipers, steering wheel, and even the crankshaft and axle housing to the rear of the car are clearly portrayed. The item which is officially missing, according to the test protocol, is the left hand front headlight. This is an obvious omission. But almost as obvious an omission, from the perspective of the street level a little in front of the car, are the back wheels. A considerable minority of children offered this latter answer.

It appeared to be a legitimate response to the question "What is missing or left out on this picture?" Although protocol did not sanction 'the back wheels' as a legitimate answer, it was marked correct but with a note identifying the children who had offered this alternative.

When the same alternative responses appeared in the second test battery (where the WPPSI Picture Completion sub-test was given again), it was decided to examine all the Picture Completion scores from 182 children in the first test battery (excluding those children who had already left the sample), as well as the scores from the first 122 children to complete the second (Reception level) test battery. This gave a total of 304 administrations of the test.

It was considered necessary to undertake a detailed analysis of the car item before making a final decision on whether to include 'back wheels' as a valid answer.

The results showed that 147 children had pointed to the missing light, 43 to the missing back wheels, and 114 had offered either 'don't know' or incorrect or unjustifiable replies such as 'the steering wheel' or 'the man driving'.

The key issue was whether the 43 children offering 'back wheels' were in fact at the end of the rising scale of difficulty, and had merely scored an unjustifiable extra point compared to the remaining 147 children identifying the missing light. Further analysis showed the following:

a. 36 of the 43 would only have lost one point if this answer was rejected. But 7 stood to lose more points further on in the test because they now had a row of five failures.

b. A comparison of scores beyond the target item showed these results:

Sample:	43 children resp. 'back wheels'	43 children resp. 'car lamp'
No further correct answer	9	6
One " " "	5	12
Two " " "	13	7
Three " " "	8	11
Four " " "	4	4
Five " " "	3	2
Six " " "	-	1
Seven " " "	1	-

It would of course be possible to analyse these figures in other ways, but it appears that there is little if any real difference. (Because of the time taken to carry out this single exercise, it was felt sufficient to base the comparison on the first 43 of the 147 'car lamp' responses.)

It was accordingly decided to include 'missing wheels' as a valid answer to this item (as an alternative to 'car lamp'), and score the sub-test accordingly.

Infant Reading Test. A study carried out by Barker (1976) in the previous research programme where this test was developed, showed that on a sample of 32 Reception class children the Infant Reading Test (I.R.T.) had a correlation of 0.89 with the Southgate Reading Test (Southgate, 1959) and 0.62 with the Daniels and Diack Sentence Test (Daniels and Diack, 1958), all three tests

being administered at the same time, at the end of the school year. Scores on the I.R.T. administered near the beginning of the school year predicted 37.5 per cent of the final variance of a combined measure of reading (using T scores from all three of the tests cited above), in a regression equation which included measures of verbal and non-verbal intelligence and of motivation; total prediction was 67.4 per cent of the final variance. (Ordinary Least Squares regression was used in that research, so that the prediction division should be treated as indicative rather than conclusive, given the small size of the sample used there.)

In the present study the I.R.T., with minor modifications to update its 'environmental' reading words, continued to perform well. Its correlation with the Southgate test, on 159 surviving sample children in the post-test attainment battery, was 0.89, while its correlation with the Daniels and Diack measure in the same battery was 0.87. These tests were administered in the Reception class at the end of the field research period.

The I.R.T. scores measured in the Nursery class, between 18 and 22 months earlier, yielded predictive validity correlations of 0.63 and 0.67 with Southgate and Daniels and Diack respectively, and 0.67 with the post-test measure of I.R.T. itself. The new test, administered on children aged between 3:9 and 4:6, remains a consistently powerful predictor of reading attainment late in the Reception year, even with the far more complex regression and path models developed for this study.

The conclusion is that the Infant Reading Test has again shown itself to be a valuable instrument for assessing 'reading' abilities in children between the ages of 3:9 and 6:0 to 6:6 years. The earlier study (Barker, 1976) presented reasons why the I.R.T. may be considered preferable to existing pre-school measures of early reading, relying as they usually do on fine motor and other skills only indirectly related to reading attainment. In contrast, the new measure assesses a range of the word and letter recognition, book handling and name writing which go to make up reading itself; the words included cover a variety of those environmentally based signs and popular words which young children tend to learn prior to learning 'school words'. It may be for this reason that it has proved a relatively powerful predictor.

6.60      Data: age and time variables

The wide range of children's ages and their differing periods in the nursery and reception classes emphasised the importance of the various age and time variables which were gathered with the other data. The sample or cohort age ranged over  $19\frac{1}{2}$  months - or 39 half months in the units of measurement employed here; testing ages also varied widely. The children differed very considerably in the range of their abilities, even within the same age levels, and there were noticeable differences in the attainment levels of the intake at different schools, according to the degree of disadvantage in the surrounding communities. The fact that cohort age covered such a wide range had a parallel effect on the age at which children were given the post-test battery. Some children had only just completed nursery class (one bright beginning reader was assessed in the nursery class prior to leaving for another city), while a few of the oldest children had already moved into a post-reception class.

Given the basic framework of regression analysis, it was possible to take account of these age and time differences by entering the temporal variables at the appropriate points in the path analysis models.

This section examines the characteristics and behaviour of the age and time variables and looks in particular at a few apparent anomalies.

Since ages were measured to the nearest half month, the question arose of whether use should be made of the standard age corrections for the WPPSI scores on the four scales employed in this project. The WPPSI standardisation tables provide norms and standard scores for each three-month age span; such adjustments are required if the scores are to be used for calculating intelligence quotients. On the other hand age adjustments cannot legitimately be made if the WPPSI scores are to be entered into multivariate regressions alongside age predictors. The WPPSI scores were thus used throughout in raw form.

There were three relatively minor problems in recording the temporal data and ad hoc adjustments had to be made. Seventeen of the sample children were found to be so immature at the initial stage that they could not be given the cognitive ability tests at the same time as the assessment of their attainment levels in reading, mathematics and Piagetian skills. Since it was expected that a number of the parents of the 17 children would be participating in the parent programmes it was not possible however to delay the assessment of their attainment levels. This minority was therefore assessed on the ability measures several months later than on the attainment measures. Since the main value of the nursery age variable was seen in its relationship to the set of ability measures, the pre-test age used for the 17 children was the age at which they underwent the cognitive tests. The effect of this disparity between the ages

of the nursery attainment and ability tests for this small group was thought to be minimal, especially as many of the immature children scored few or no points on most of the attainment variables.

Illness and unexpected absences also meant that a few other children were not given the attainment and ability batteries in the same month, although here the gap in test ages was usually only about a month, depending on E's test schedule.

A third problem concerned the fact that several of the schools introduced full-time nursery schooling towards the end of the nursery period. Thus a number of children were switched from morning or afternoon sessions to full day sessions. In such cases it was decided, after examining the classroom situations and the activities offered to these youngsters, to assess the value of the full day attendance as one and two-thirds the value of the half day sessions. This was judging the experience in terms of its estimated educational impact. In terms of social impact - had that been the project's focus of attention - one would probably estimate the value of the full day attendance to be at least two and a half to three times the value of a single session. The latter judgement was based on the added value of the ordered lunch-time sessions in which the children were guided into waiting their turn and sitting down to formal meals, with the nursery teacher, nursery assistants and dinner ladies emphasising basic behavioural requirements.

#### 6.61 Age relationships

The histograms of the three test ages, Nursery age, Reception age and Post-test age, appear in figure 17.

Table 30 presents a selection of the more important correlations between test ages and some of the attainment and ability variables. For each test age the relationships with the variables assessed at that particular age are enclosed in a box. This is to set them out from the correlations between the test ages and the variables assessed at other ages.

The comparison of the correlations across the different test ages was undertaken to examine the seemingly unusual and rather questionable relationships between test ages and some of the cognitive test scores. As already mentioned, the scores for the four WPPSI ability tests were not corrected for the age of testing, in view of the fact that test ages were to be entered at several points in the path analysis. This makes it all the more surprising that the WPPSI scores show low and even a few negative relationships with age.

The correlations of different test ages with the same test scores indicate

Figure 17. Histograms of Nursery age, Reception (mid-test) age and Post-test age.

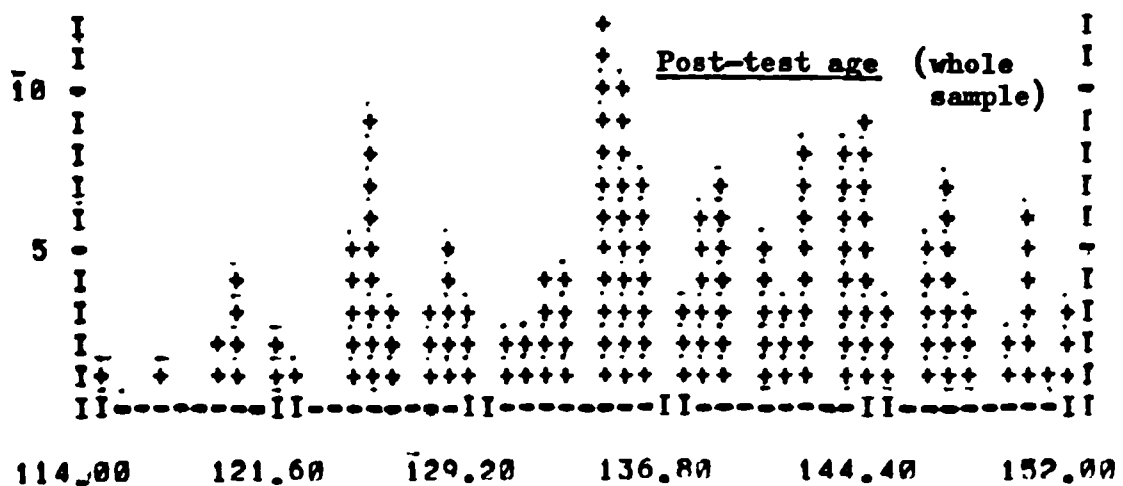
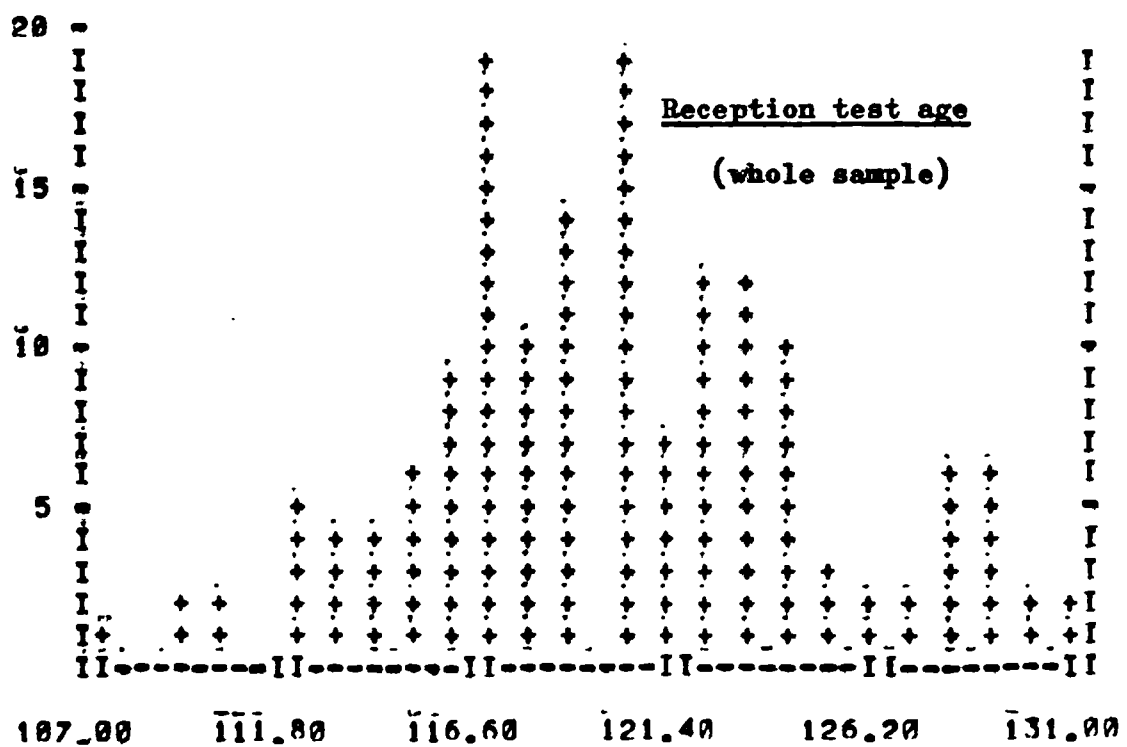
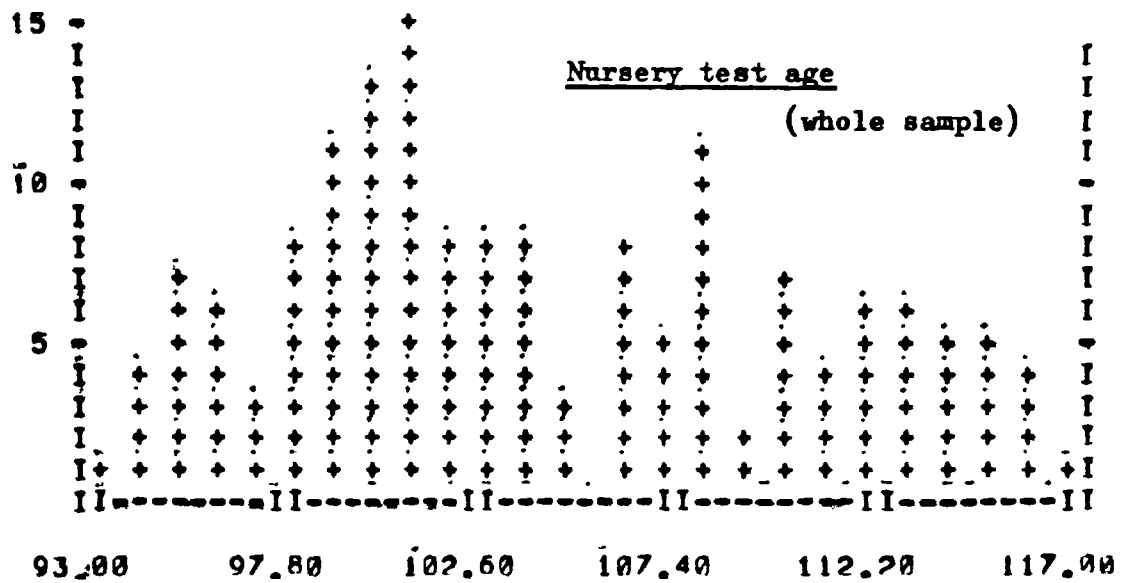




Table 30. Selected correlations between age, ability and attainment variables

	<u>Corr. with Nursery age</u>			<u>Corr. with Receptn age</u>			<u>Corr. w. Post-Test Age</u>
	Nursery battery	Receptn battery	Post-T. battery	Nursery battery	Receptn battery	Post-T. battery	Post-T. battery
<u>Ability:</u>							
Information	.010	-.134		-.043	-.102		
Sentences	.055	-.050		.014	.050		
Pic.Complet.	.147	.080		.062	.005		
Block Design	.239	.107		.038	.021		
Bend.Gestalt	.213	.100		.049	.127		
<u>Attainment skills:</u>							
Inf.Reading	.186		.307	.102		.295	.388
Maths Numer.	.156		.190	.084		.120	.234
<u>Conceptual attainment:</u>							
Maths concep.	.013		.040	.002		.069	.089
Piagetian	-.067		.175*	-.069	-.220	.146*	.223*
Test age	1.000	.744	.890	.744	1.000	.833	1.000

\* These particular correlations apply to the Piagetian test scored according to the successive failure protocol, as described in appendix A3, test 9. The comparative correlations when the Piagetian post-test results are scored according to the original protocol are: -.106 (with Nursery age); .144 (with Mid-test age); and .160 (with Post-test age).

that these results are genuine and not merely due to error in measuring one of the age variables. Nursery age correlations with the tests carried out in the reception battery bear a reasonable relationship to the set of correlations between nursery age and the same variables measured earlier in the nursery battery, given the fact that the age distribution for the reception battery was much more leptokurtic or peaked than that of the nursery battery. Likewise the reception age correlations across the nursery and reception battery variables bear a reasonable relationship to each other, given the known variability in correlations and the different age structures for the two batteries. It is noticeable that the correlations with reception age are on the whole much lower than those of either nursery age or post-test age. It is reasonable to contend that this may be related to the more leptokurtic nature of the reception age distribution.

While the above may be a credible explanation for the variations in the relationships across test ages and test batteries, it emphasises the danger of

relying too much on the evidence of single correlations. Before dealing further with the possible reasons for the anomalous relationships between age and the ability variables it is necessary to review the possibility either that the age variables have been erroneously measured, or that the test scores have little or no validity.

The mechanical method employed to calculate the age scores offers little chance that gross errors could have been made in the scoring, which was estimated on the instruments to an accuracy of within less than a week, with holiday periods excluded. A cross check was used in which the different age scores were combined for each child to yield a constant. The detailed examination of the reliability and validity of the test battery variables, in section 6.50, likewise leaves little room for the possibility that the variables cited in the above table have no real validity.

What can be advanced instead is the proposition that the nature of the cohort and the fact that the testing was carried out sequentially in six different schools together offer a coherent alternative explanation for the anomalous results. There are several points that can be noted.

Firstly, the nature of the nursery sample was such that the children in the advantaged school tended to be both younger and brighter than most of the children in the five disadvantaged schools. Even within the advantaged school a handful of unusually bright children were among the youngest in that school's sample. The question of these differing school intakes has already been dealt with at some length in section 6.40. Such factors contribute to a reverse age effect in which a normal modest but positive relationship between test scores and test age is dampened or even occasionally reversed by the nature of the cohort. This problem can be dealt with in closely controlled experiments by administering the tests to each child only when he or she is within a week or two of some uniform target age, or alternatively by ensuring a homogenous random sample from the population. Such controls would not be possible in field studies such as the present one.

Secondly, the nature of the reception age distribution merits attention. It is much closer to normal than either the nursery or post-test ages. This is because of a relative though not absolute uniformity in the age of transfer from nursery to reception class. The closer uniformity in the age of testing, as evidenced by the more leptokurtic distribution, would naturally tend to show a lower relationship between age and variables which are to some extent age dependent (in the extreme case of testing all children at the same age the correlation with age would clearly be zero).

Thirdly, it should be noted that these anomalous effects are most marked in the measurements of the ability tests, all of which have a relatively limited

range of scores and have been designed to cover a wide age range (in the case of the WPPSI tests, from the age of 3:10 to 6:7 years). In contrast the reading and mathematics tests used in this research were designed to have wide score ranges and to assess attainment during the last year of pre-school and the first year of formal school. Thus the age relationships with the attainment tests are never very low; it can however be noted that even here there is a suggestion that the reverse age effect due to the sample composition is operating to reduce correlations slightly.

Of the two variables labelled 'conceptual attainment', mathematical concepts has a consistently low relationship to age, even to post-test age. This suggests that the attainment measured is not closely related to age but is possibly more susceptible to instruction. The Piagetian test variable shows a negative relationship with both nursery and reception ages. While the initial negative correlation with nursery age can be interpreted as a reverse age effect in a variable that may well be influenced by experience, the high correlation with reception age of  $-.220$  is too large to be interpreted in this way; its two-tailed probability level of around  $.02$  within a table of 48 correlations does permit recourse to the explanation of chance - although in fact one-tailed relationships in the positive direction were all that were predicted here.

Another possible explanation might be that during the reception tests a number of children - by now much more confident than at the time of the nursery tests - offered what they thought were the 'smart' answers to several Piagetian tasks and this may have served to worsen a correlation already subject to a sample reverse age effect.

There remains one major question. If the explanation of a biased distribution of the relationships between test ages and test scores is to be accepted, does this imply that age variables should be eliminated from the analyses? This raises again the issue dealt with in section 5.11 on the limitations of correlational evidence.

Correlations can give a very different picture of a relationship than regressions do, despite the fact that regression coefficients are derived from an inverse correlation matrix. Multiple regression equations computed in this study, and presented in section 6.80, show that the age variables contribute substantially to the post-test criteria, even alongside batteries of the other predictor variables. Moreover, the study is focused on the prediction of post-test academic attainment; the age variables have a consistently strong and positive correlation with pre and post-test attainment variables, in contrast to the problematical relationship with some of the ability variables.

Finally, regression is a robust technique, particularly when relying on the V-ridge regression procedures developed in this study; awkward distribu-

tions in predictor variables are no barrier to the use of regression, nor can their awkward relationships with other predictors be regarded as a hindrance, given the overall power of the inverse matrix of correlations between all the predictors, with the V-ridge constant serving to stabilise and 'smooth' the matrix of relationships.

#### 6.62 Time relationships

The nature of the two time variables, 'time in nursery' and 'time in reception', have been discussed briefly in section 5.17. These variables are measures of actual half months spent in nursery or reception, with holiday periods excluded from the totals but with a compensating increase for full-day nursery attendance, as explained in an earlier section. The third variable intimately involved with these two is cohort age, a measure of the 'distance' between 1 January 1972 and the subsequent birth dates of the sample children; thus the oldest children have the lowest scores, since their births are the closest to the datum line.

During the examination of these variables a variety of approaches were studied in an attempt to determine how to assess the effects of time in nursery and time in reception on the children's performance at the end of the study period, taking into account the children's ages relative to the rest of the cohort sample.

Figure 18 presents the histograms of cohort age, time in nursery and time in reception, prior to the adjustment of the time variables.

One major approach, to which reference was made in section 5.17, was the attempt to regress three of the post-test attainment variables on to cohort age and time in reception jointly. The criterion variables were, in turn, the composite reading variable, the composite maths concepts variable and maths numeracy. Prior to carrying out the regressions the variables were all school-centred - in other words, each child's scores were centred on the means for her or his school. This was done to eliminate the effects of slight differences in transfer policies across the different schools.

Table 31 sets out key results from the three regression runs. Clearly both cohort age and length in reception contribute a fair amount to composite reading, at least in the absence of other competing variables. For mathematical concepts cohort age makes a modest contribution, but the contribution of time in

Histograms...../

Figure 18. Histograms of raw cohort age, time in nursery and time in reception.

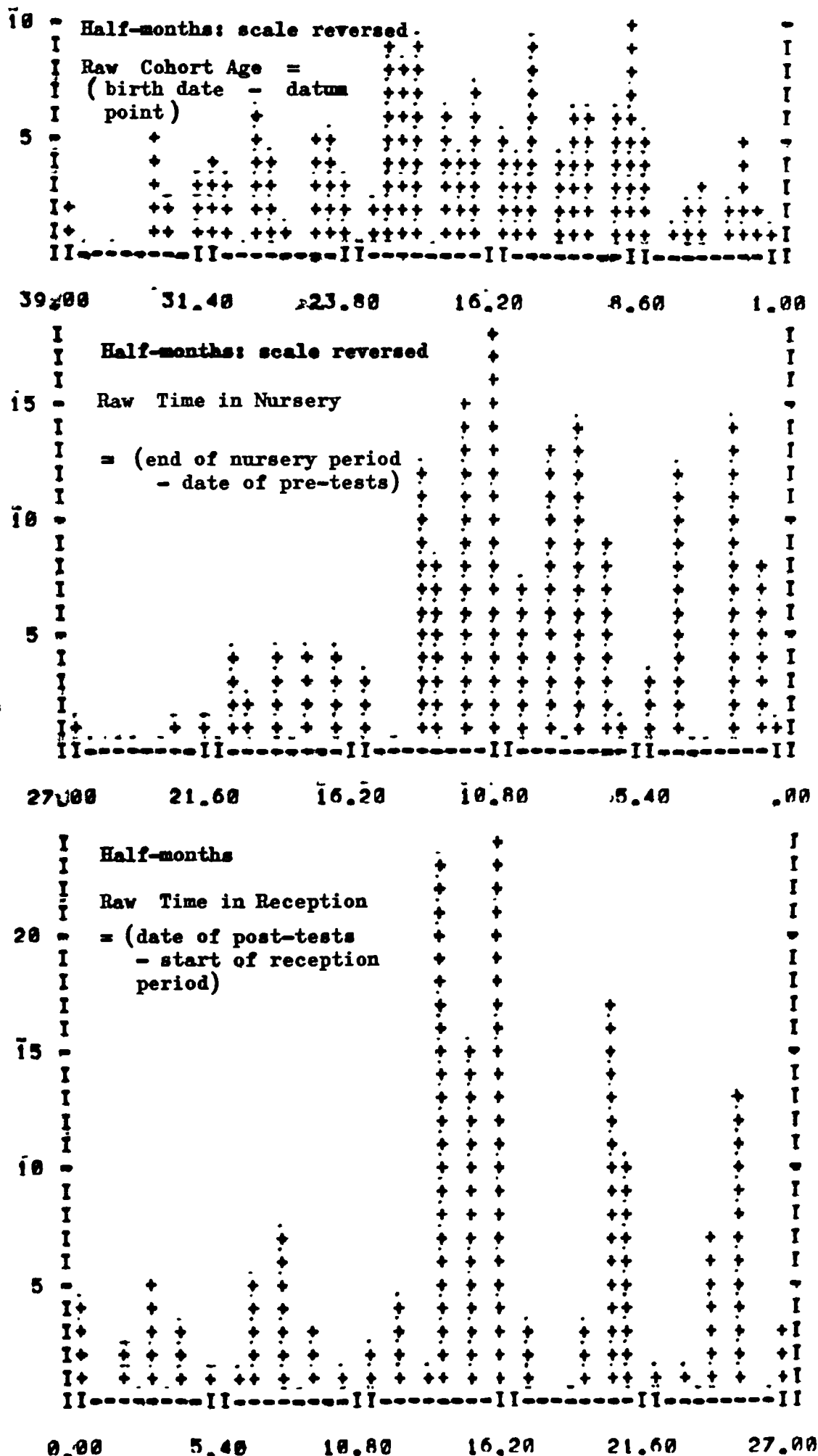


Table 31. Results of attainment regressions on cohort age and time in reception, using whole sample (N=159)

	Dependent variable					
	Post-test composite reading		Post-test mathemat. concepts		Post-test mathemat. numeracy	
	Coefft. and p*	Unique vaf (%)	Coefft. and p*	Unique vaf (%)	Coefft. and p*	Unique vaf (%)
<u>Predictors:</u>						
Cohort age	.241 (.0000)	10.24	.157 (.0008)	2.56	.110 (.011)	1.12
Time in reception	.150 (.0004)	4.56	.049 (.15)	0.28	.157 (.0004)	2.62
<u>Variance predicted:</u>						
Shared vaf (%)		2.64		0.56		2.98
Total vaf (%)		17.43		3.40		6.58

vaf Variance accounted for (percentage of total variance of dependent var.)

reception is too small to be acceptable. A higher proportion of mathematical numeracy is predicted by the two variables, with time in reception being more important than cohort age.

It is possible to contend that these results are intuitively acceptable. Neither the parents nor the curriculum in the nursery classes had introduced children to numeracy or manipulation of numbers, so that it could be expected that time in reception should have a larger effect on performance than does cohort age. On the other hand reading had long been introduced to the children, either through the parents reading stories to them or even teaching them to read some words, or through listening to stories in the nursery class. Thus it is credible that cohort age - to some extent a reflection of length of experience - should contribute heavily to reading performance. However time in reception, equivalent to teaching time, also contributes considerably to reading levels.

The fact that time in reception appears to make little contribution to mathematical concepts is interesting; only cohort age makes any meaningful addition to predicted variance and even that is small. This suggests that some other kinds of variables are mainly responsible for the development of these concepts. The issue is discussed in detail in section 6.80.

It can be noted that the centering of these variables on the school means

has resulted in a small increase in the total variance predicted and, of more importance, it alters or even reverses the relative size of the contributions of the time and cohort age variables in some cases.

The variable time in nursery (with scores reversed) shows much the same pattern of relationships to cohort age and post-test outcomes as does time in reception. This is to be expected, as time in nursery and time in reception are inversely related.

It is however difficult to separate the relative effects of these two highly correlated time variables, although the process of entering each time variable at appropriate points in the path analyses could enable the contributions to be assessed within a global model in which each 'latent variable' is in competition with all the other contributors acting at that point in time. (The strength of the inverse relationship between time in reception and time in nursery is indicated by their correlation of 0.936, with time in nursery scored negatively.)

It is even more difficult to interpret the two time variables as entities distinct from a simple age variable, since even with the joint regressions (with cohort age as a co-predictor) it is clear that both the age and time variables are competing for variance on the strength of their powerful links with age of testing (time in reception having an indirect positive link and time in nursery an indirect negative link). It was therefore decided to abandon this approach in favour of taking account of cohort age within each of the time variables.

#### 6.63 Derivation of new time variables

As the exact nature of the two time variables, when standardised on the school samples and adjusted for cohort age, is of particular interest in the path analyses, it is worth offering a more detailed discussion of their nature at this point. To distinguish between the raw time variables and the derived ones, the former will now be termed Length in Nursery and Length in Reception, and the latter Time in Nursery and Time in Reception.

The new time variables are defined as follows:

$$\text{Time in Nursery} = \text{School-standardised Length in Nursery} - \text{School-standardised Cohort Age}$$

$$\text{Time in Reception} = \text{School-standardised Length in Reception} - \text{School-standardised Cohort Age}$$

The diagram on the following page illustrates the interrelationship between the different periods for two typical children.

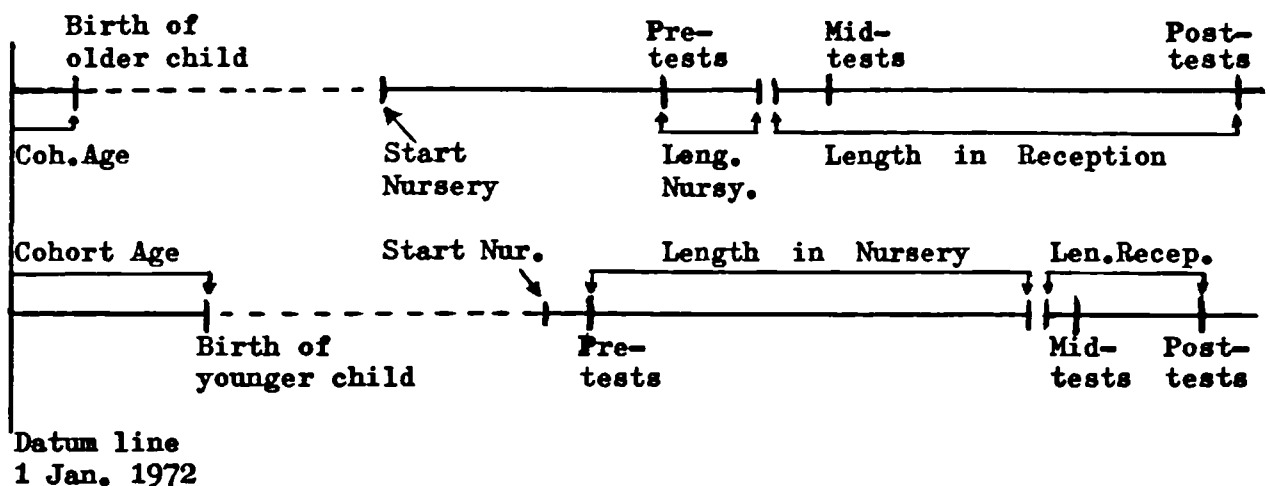


Figure 19. Diagrammatic representation of cohort age and time concepts

Although the composite variables appear to be relatively simple in their construction there are some problematical issues. There is clearly a direct relationship between Length in Nursery and Cohort Age, and an inverse relationship between Length in Reception and Cohort Age. From this point of view it might therefore be argued that Cohort Age should indeed be subtracted from Length in Nursery but should on the other hand be added to Length in Reception. In theory this should eliminate age effects and leave any remaining variance in the two combined variables to yield whatever predictive power they had in the multiple regression equations alongside all the other variables (including the age of assessment variables).

However the above consideration does not take into account the effects of the standardisation of each of the variables. A different way of approaching the concepts is to examine the situation of three children, one younger than the average, one of average age and one older. Their standardised Cohort Age scores might be 1.0, 0.0 and -1.0 (since older children have smaller raw cohort age scores). If these three children had all spent less time in the Nursery (between time of testing and time of leaving for Reception) than the mean sample time - say -1.0 on a standardised scale - then their Time in Nursery scores would be as follows:

The younger child	-1.0	-	1.0	=	-2.0
The average age child	-1.0	-	0.0	=	-1.0
The older child	-1.0	-	-1.0	=	0.0

These scores are as expected, namely that the younger the child the smaller the variable score for Time in Nursery. In other words for children with the same nursery experience the age difference is directly reflected in the variable score. It may be argued that this hardly appears to be an example of the removal of age from the time variable. What has happened in fact is a



correction of the length scores to take account of the age effect; while the score of -1.0 may appear to credit the three children with the same Nursery experience, in fact their differing ages would have to be taken into account in judging the predictive strength of this Time variable. (The alternative strategy is the one mentioned in the previous section, where length in nursery and cohort age are entered in a regression as joint predictors.)

In contrast to the above example, differences in the period spent in the nursery for children of the same age are directly reflected in the composite measure. For example, if three children with ages equally above the mean age, say +1.0 on a standard scale, had spent differing periods in the nursery, their scores on the time variable would be as follows:

Shorter than average Time in Nursery	-1.0	-	1.0	=	-2.0
Average Time in Nursery	0.0	-	1.0	=	-1.0
Longer than average Time in Nursery	+1.0	-	1.0	=	0.0

Exactly the same principle applies to the determination of the Time in Reception variable. Differences in the children's ages for the same period in reception are reflected in the variable scores, as are differences in time spent in reception.

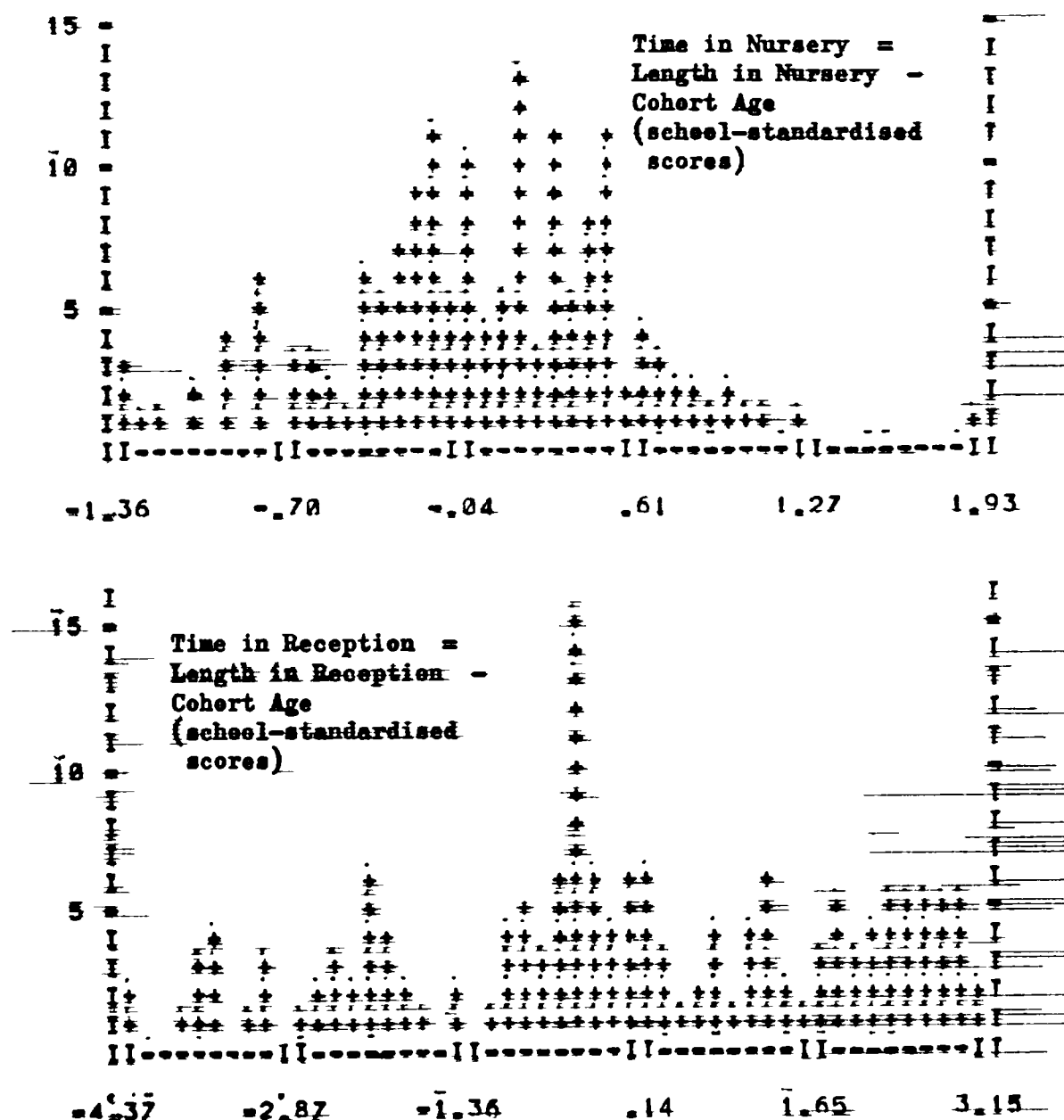
There appears to be a conceptual problem arising from the fact that children of a lower cohort age (older children) tend to have a longer period in reception, while the reverse is true for the period in the nursery. The subtraction appears to cancel out two parallel scores in the case of Length in Nursery and Cohort Age, while the same subtraction appears to exaggerate the opposite polarity of Length in Reception and Cohort Age. This conceptual issue does not however of itself vitiate the validity of the two newly created variables. Their characteristics and their relationships with variables such as Age at Post-Test are sufficiently similar to suggest that there is no fundamental difference in the nature of the two Time measures.

It has to be recognised however that both the new variables are still strongly suffused with age, as shown in the simple examples above. The advantage they have over the original Length variables is that they have been corrected for differences in the standardised ages of the children within each school sample. This avoids the misinterpretations which might arise from forming composite scores for the whole sample regardless of the small but possibly important differences in school policies regarding transfer from Nursery to Reception.

Histograms of the two new Time variables are shown in Figure 20 overleaf.

In conclusion, the following points should be noted in assessing the reality or otherwise of the Time in Nursery and Time in Reception variables.

Figure 20. Histograms of composite Time in Nursery and Time in Reception



a. The regressions reported in the previous sub-section offer support for the view that Length in Reception and Length in Nursery are identifiably separate predictors from the major age variable of Cohort Age. However the rather differing predictive strengths of Time in Reception and Time in Nursery (in the path models yet to be presented) suggest that the standardisation of the variables and the subtraction of Cohort Age has radically altered their character.

b. Both the new Time variables still have a strong underlay of age in their make-up, although this can be justified by the need to take into account differences in school-standardised ages across the sample.

c. At every level in the path models age variables are entered into the regression equation, competing thus with the age component of the Time variables.

At the first two levels age variables predict indirectly through their contributions to the Ability latent variables. At the third ((Post-Test) level Age at Post-Test is entered directly alongside Time in Nursery and Time in Reception, so that additional unique contributions from the two Time variables can be interpreted as genuine additions to variance explained.

d. Age at Post-Test is in theory the most powerful of all the age variables, since it reflects the exact age at which the Post-Tests were administered. Yet in three of the ten path models the Time variables have stronger relationships with the outcome variable than does Age at Post-Test. This could not easily occur if the Time variables were simply weaker surrogate age measures.

e. In eight of the ten models Time in Reception has a stronger relationship with the outcome variable than does Time in Nursery. This is to be expected, given the more clearly academic curriculum of the Reception class. The fact that Age at Post-Test is a directly competing variable in the final regression equations serves to limit the predictive strength of the two Time variables at this level. The apparently higher predictive strength of Time in Nursery at the previous level may be partly due to the fact that the Nursery and Reception variables make their contribution indirectly at this level, through latent variables.

Thus, while a certain caution is needed in interpreting the value of Time in Nursery and Time in Reception, there is enough evidence in the path models to suggest that these variables do confirm the contribution of both the nursery and reception curricula to academic performance by the children, even in competition with the many other variables entered into the models.

## 6.70 Correlations and Group Scores

Tukey (1954) wrote in light vein that he had been an early member of an informal society for the suppression of correlation coefficients. He considered that many perplexities and confusions resulted from the use of such coefficients. They were justified in only two circumstances - for use in (the derivation of) regression coefficients, or when the measurement of one or both variables on a determinate scale was impossible. Tukey cites Karl Pearson's interpretation of genetic models with the aid of correlation coefficients as one of the few examples where this statistic could be of real value, since the similarity of the standard deviations of characteristics across the generations (within the same sex) meant that correlations served almost the same function as regression coefficients.

Recent developments in genetic theory, as set out by Eaves et al (1977), put even that usage into doubt since the authors of the latter study show that critical differences in standard deviations between groups demand the more sophisticated handling of regression models to interpret the effects of genotype interaction and covariation. As Eaves et al point out, "Correlations are only an effective starting point for an analysis of individual differences when the causes of (such) differences are fairly simple".

The 1954 review by Tukey also comments on the difference between the anti-theoretical approach which relies on quantitative description and correlations within an individual experimental sample, and the wider and more theoretically challenging approach based on both quantitative and concealed elements within causal models founded on regressions. Among the other reasons advanced by Tukey in favour of regression analysis is its potential stability; by comparison correlations cannot remain the same over a wide range of situations.

Even today a great many research findings are based on the rather tenuous evidence of primary (univariate) descriptive data such as means and differences and on the secondary bivariate relationships presented by correlations and contingency tables. Although the insights afforded by these measures and derivations are of real value in the preliminary stages of any analysis they are often limiting and potentially misleading. A few examples from the present study may illustrate these dangers.

a. Although by deliberate design no parent was able to attend both mathematics and reading programme meetings - and of course a considerable minority attended neither - the correlation coefficient between mathematics meeting attendance and reading meeting attendance was a significant  $-0.37$ , suggesting that

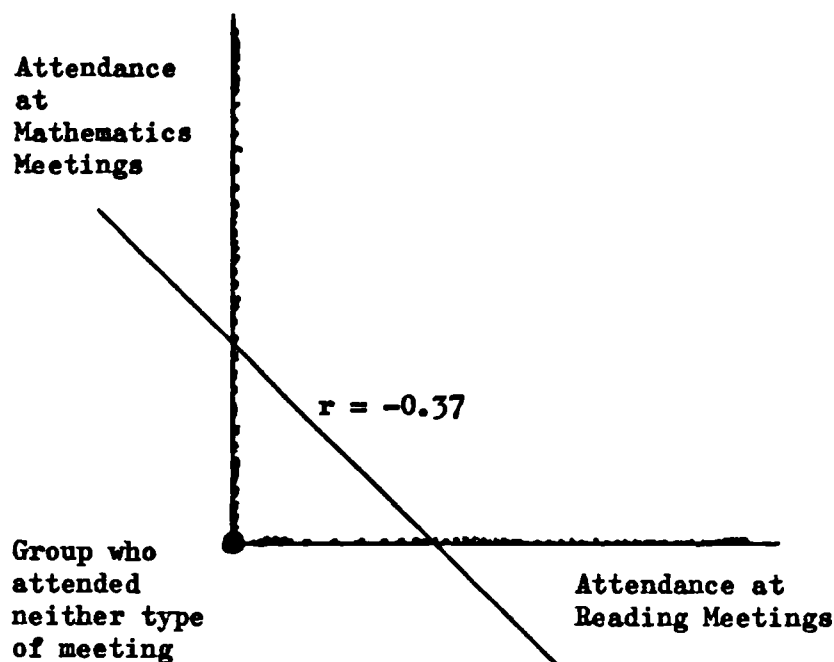


Figure 21. Plot of relationships between attendance at Programme Meetings

the more meetings of one type attended by parents, the less meetings they would attend of the other type. An examination of the figure above illustrates why it is possible to derive a sizable coefficient in a situation where one would expect a zero relationship.

b. In the action research situation where parents were invited to attend programme meetings, a number declined on the grounds that they were at work in the day-time. Although the children of these working parents (where the families were assumed to be upward mobile) were found to perform better than the children of 'stay at home' parents who attended the programmes, there was some hope that the gap might have been reduced as a result of the programmes. This hope was not realised. However it was also found that there was half a standard deviation in age levels between the children in the two groups - the Working Group's being older than the Programme Parents' children - so that no real comparison could be made between them. There was also some evidence that many Programme parents planned to go out to work as soon as the child was at or close to the age of starting full-time school; thus the assumption of upward mobility may have been only partially correct (there was some evidence pointing in that direction), while the alternative explanation that parents of older nursery children were more inclined to go out to work could equally well fit the situation and also explain the differences in achievement levels.

c. While the mean age gap between the advantaged and disadvantaged samples was only 0.8 half months for the Post-Test battery, the gap had been 2.0 and 2.7 at the time of the Nursery and Mid-Test batteries; the gap of 2.7 half months (the units used for greater accuracy) was more than half a standard deviation. Given that some of the age variables predicted more than 8 per cent

of the outcome variance in certain multiple regression equations, the enlargement or contraction of age differences clearly made direct comparisons of means even more obscure. Chance differences between the mean ages of the reading and mathematics programme children had also altered somewhat by the time of the Post-Test, due to the complex nature of testing 159 children in six different Metropolitan reception classes and in 20 other schools where 23 of the sample children were traced after they had left the original six schools.

d. Despite a conscious effort to allocate programme parents randomly into reading and mathematics groups, though limited by the need to make individual groups as complete as possible within any one school, there were considerable differences in both child ability and parent behaviour variables, making it impossible to treat these groups as completely equivalent or comparable.

e. Although the English Picture Vocabulary Test scores correlate highly with the reading outcomes in all the models, and would appear therefore to be a valuable predictor of Post-Test attainment, within the multiple regression path models the E.P.V.T. makes a zero contribution in almost every sample group. It does appear as a strong predictor in one sample, however, and the contrast between this and other samples may be of considerable research interest. Only a multivariate model could have detected the massive differences in prediction which are brought out by the path models in section 6.80.

Difficulties such as those outlined above point to the major limitations of simple analyses. The intense debates which have taken place in the American research literature on issues such as that of school effectiveness (Coleman, 1966, and the subsequent national debate leading up to Jencks, 1972, et seq.), claimed ethnic differences in ability (Jensen, 1969, followed by a considerable volume of academic re-analysis and political argument), and compensatory education (the critical Westinghouse/Ohio analysis - Cicirelli et al, 1969 - and a great deal of subsequent debate and analysis culminating in the recent major review by Lazar et al, 1977) have all been clouded if not confused by arguments over means and correlations.

There are few if any major explanatory models, based on multivariate analysis, which cannot be re-interpreted in the opposite direction by a narrow focusing on some primary descriptive data or simple bivariate relationships. Because the latter are more easily grasped they are more open to misuse; equally the counter-arguments based on advanced analyses are more difficult to comprehend, let alone to present.

While the preceding paragraphs have spelt out the caveats over reliance on evidence based simply on univariate or bivariate statistical data, there

are some interesting pointers that can be drawn from such statistics, tentative though they may be. The following sections outline some of these possibilities.

#### 6.71 Transformations of variables

A number of variables were subjected to transformations with a view to improving their performance in the path models.

English Picture Vocabulary Test. Although this variable showed only a moderate skew, its poor predictive performance in most of the models suggested that this might be related to its distributional characteristics. Consequently the square root, log and squared transformations were derived. The changes in its correlation with the major outcome of Infant Reading Test were so minimal that there appeared to be little justification for introducing the transformed variable into the path models, based as they are on standardised data. (It was recognised that regression relationships might be altered more than the correlations were, when transformations were made in a predictor variable. However the implications of transforming one predictor and not the others within a multiple regression situation would have required extensive experimental study in order to justify such changes.)

Other pre-test variables. Several other pre-test variables whose performances were not particularly strong in the multiple regression equations - such as Distractibility and Nursery Needs - were also subjected to various transformations, mainly the log and square root. Again the changes in correlations with a range of outcome variables were not such as to justify any further work on the transformed variables.

Skewed post-test (dependent) variables. While the Infant Reading Test yielded a valid score for every child in the Post-Test battery, both the Southgate Reading Test and the Daniels and Diack Sentences Test had a number of zero scores. This was particularly serious for the Daniels and Diack test, where nearly half the sample failed to read any sentences. For Southgate the number of zero scores was less than one-sixth of the total. On the other hand both these tests served as discriminators for high performing readers who on occasion reached ceiling or close to ceiling in the Infant Reading Test. The scores for Southgate and Daniels and Diack were accordingly normalised and their correlations with the I.R.T. and other variables were examined. It was finally decided that, given the weighting of the Post-Test reading outcome, which is

$$\text{Post-Test Reading Attainment} = (0.4 * \text{I.R.T.}) + (0.4 * \text{Southgate}) + (0.2 * \text{D.Diack}),$$

using standardised scores in each case, there was little reason for adding to the complexity by normalisation of the components prior to standardisation and formation of the composite variable, especially as this normalisation would have had to be repeated for each separate sample in the regression models. The combination of low and high-scoring reading measures clearly reduced the size of the skew and appeared to yield a respectable distribution for the total sample.

## 6.72 Group scores

For reasons set out at the beginning of this section it would be unwise to place too much reliance on the size of the sample mean scores or on the differences between them, especially if one takes into account the differences in mean age levels and the presence of other intervening variables. As already explained, such contamination necessitates a multivariate rather than a simpler approach to the results of this study. For the same reason no t-tests will be performed on sample differences, as such tests assume a fundamental equality of samples in every respect other than that of the variable under scrutiny.

A small selection of group mean scores is set out in the table overleaf.

There are other forms of comparison which can be made from the early stages of data analysis within the regression models.

Girls versus Boys. Within the Disadvantaged sample, girls and boys scored approximately equally on the pre-test Infant Reading Test, but girls were ahead of boys on the pre-test Maths and Piagetian measures (differences ranging from a quarter to two-thirds of a standard deviation). On a composite measure of Post-Test reading and mathematics scores (to be described in full in section 6.80), girls scored approximately one-fifth of a standard deviation ahead of boys.

Blacks (West Indians and Africans) versus Whites (English and Irish). Here the differences (within the Disadvantaged sample) were found to favour the Whites in both the Pre-Test and Post-Test attainment measures. On the composite Post-Test measure described in the previous paragraph, White children scored a quarter standard deviation ahead of Black children. The environmental factors thought to contribute to these differences are discussed in section 6.424 and in model 9 in section 6.80.

Working Group versus Programme Attenders. Within the Disadvantaged sample, children of the Working Group parents (who by definition could not attend programme meetings) scored nearly half a standard deviation ahead of the child-



**Table 32 . Selection of group mean scores and standard deviations for various samples**

Variable	Whole Sample	Advantaged School Sample	Disadvantgd. Schools Sample	Disadvantgd. Reading Programme Attenders	Disadvantgd. Maths Programme Attenders
Pre-Test Infant Reading Test	5.92 (6.08)	8.33 (7.60)	5.43 (5.66)	4.75 (4.47)	5.25 (4.68)
Pre-Test Maths Numeracy	5.93 (2.47)	6.82 (2.17)	5.77 (2.48)	5.65 (2.39)	5.84 (2.19)
Pre-Test Maths Concepts	4.38 (1.93)	5.19 (1.50)	4.25 (1.98)	4.10 (1.98)	4.75 (1.80)
Pre-Test Piagetian Tests	1.36 (1.56)	2.04 (1.93)	1.23 (1.45)	1.33 (1.52)	0.97 (1.23)
Post-Test Infant Reading Test	28.09 (14.84)	29.74 (16.20)	27.98 (14.64)	27.75 (13.17)	27.72 (13.67)
Post-Test Southgate Reading Test	8.77 (6.73)	8.67 (7.84)	8.88 (6.53)	8.71 (6.07)	8.88 (5.79)
Post-Test Dan. Diack Reading Test	9.23 (14.31)	9.74 (16.37)	9.21 (14.02)	7.60 (11.93)	8.03 (12.33)
Post-Test Maths Numeracy	11.28 (2.89)	12.78 (2.59)	10.97 (2.89)	10.73 (2.92)	10.72 (2.72)
Post-Test Maths Concepts	6.90 (1.86)	7.96 (1.61)	6.70 (1.85)	6.63 (2.02)	6.88 (1.64)
Post-Test Piagetian Tests	6.38 (3.51)	8.30 (3.70)	5.97 (3.38)	5.44 (3.24)	6.22 (3.13)

ren of Programme Attenders in the Pre-Test reading measure, but only fractionally ahead in two of three other Pre-Test attainments. For the Post-Test composite measure, Working Group children scored approximately one-third of a standard deviation ahead of the children of Programme Attenders.

Non-Attenders. The non-working Non-Attender group in the Disadvantaged Sample (21 dyads) scored consistently below the Programme Attenders group, in both the Pre-Test and Post-Test attainment measures. The gap widened from less than one-tenth of a standard deviation in the Pre-Test set to one-sixth in the Post-Test composite measure.

School differences. For several reasons it was felt inappropriate to compare differences across schools at any level beyond a recognition of the persisting and well known differences between advantaged and disadvantaged schools. At the outset of the study, in the negotiations with the Metropolitan Local Education Authority, it was stated by E that the purpose of the study was not to examine the schools as such nor to compare the performance of teachers within or between schools. For any competent comparison it would have been necessary to obtain detailed socio-economic and other relevant data from the school catchment areas, as well as a variety of school data such as ethos and curriculum descriptor variables. Even then an examination of means or other sample differences would have been wholly inappropriate, for the reasons set out earlier in this section. Possibly the most valid form of comparison between schools would be in the form of a separate path model for each school, using a range of parent, child and school variables and grouped variables as input, and examining the relative contributions of each of these sources to child attainment at some fixed class or age level.

The disparities in ages between the various samples, and the widening of some of these age gaps during the course of the study, have already been discussed. Even had the age differences not been present it would have been highly questionable to base any firm conclusions on apparent group differences, given the wide range of variables which contribute to performance within each group. The fundamental methodological question to be asked and evaluated in the subsequent major analyses is whether the statistical model employed here, and its conceptual foundations, are adequate and credible as evidence in favour of the hypothesis that parent programmes during the Nursery period are of use in stimulating performance by the children concerned. The ultimate goal is the examination of the educational hypothesis that the parent programmes have significant and meaningful effects on children's later performance in reading and/or in mathematics.

#### 6.73 Predictive power of meetings attended

The four variables of most relevance to the research hypothesis are those of meetings attended. Two of the variables (scored separately for reading and mathematics programmes) are direct measures of the number of meetings attended by each parent; the other two variables provide attendance scores which have been weighted to take account of the amount of child 'disturbance' occurring at each meeting. These measures are entered as 'Parent Programme' variables in each of the main path models (apart from the model for Working Group parents who did not attend any meetings). While the validity of the measures of

meetings attended, and their possible contamination by other (prior) parent variables, will be closely examined in section 6.80, a useful preliminary correlation and regression exercise examined whether in fact attendance at reading meetings predicted higher post-test mathematics scores, and equally whether attendance at mathematics meetings predicted higher post-test reading scores.

A stringent test was undertaken to examine the possibility of cross-prediction. Scores were used from the whole sample, thus considerably reducing the predictive power of the meetings variables, as less than 40 per cent of the sample (60 out of 159 parents) attended any reading meetings and only 25 per cent (39 parents) attended any maths meetings. (As explained earlier, attendance at one type of meeting precluded attendance at the other type.)

The table below presents the correlations of the four meetings variables with the composite score of Post-Test Reading attainment, and (separately) with the composite measure of Post-Test Mathematics Concepts. The results of regressing each of these two Post-Test scores simultaneously on all four meetings variables are also presented in the form of regression coefficients and probability data. Since the whole sample is used, rather than isolating the reading and mathematics programme groups, the variance predicted in each case disappears once the statistical adjustments have been made for the numbers of cases and variables.

Table 33. Correlations and regression parameters for Programme meetings

Variable	Correlation with Dependent var.	Regression coefficient	Probability of regr. coefft.
<b>a. Dependent variable: <u>Post-Test Reading</u> (composite)</b>			
Reading meetings	0.1140	0.049	.068
Reading meetings (weighted)	0.1047	0.029	.239
Mathematics meetings	-0.0226	0.007	.438
Mathematics meetings (weighted)	-0.0234	0.004	.463
<b>b. Dependent variable: <u>Post-Test Mathematics Concepts</u> (composite)</b>			
Reading meetings	0.0315	.008	.390
Reading meetings (weighted)	0.0380	.038	.089
Mathematics meetings	0.0891	.021	.293
Mathematics meetings (weighted)	0.1027	.066	.045

It should be emphasised that these predictions are in isolation, in contrast to the principle observed throughout the path models where each variable has to prove its predictive power in competition with the other relevant predictors. However the stringency of this method, using all 159 cases together, offers important evidence that there is no inevitable correlation between mere attendance at any meetings whatever (which would suggest a heavily biased sample of parents) and improved performance in the Post-Test attainment measures.

In reality, the table on the previous page offers strong suggestive evidence that it is specifically attendance at reading meetings rather than at mathematics meetings which predicts to Post-Test Reading attainment. The correlations between attendance at mathematics meetings and reading attainment are in fact slightly negative.

While the evidence on attendance at mathematics meetings and on the prediction of Post-Test Maths Concepts is not as clearcut as in the reverse situation, the differences are still sufficiently large to claim that it is the weighted mathematics meeting variable - rather than either of the reading meeting variables - which predicts significantly to the Maths Concepts outcome.

(Throughout the path analyses it has been found that the weighted mathematics meeting variable predicts better than the unweighted measure, suggesting that for these meetings the amount of child disturbance at the meetings did affect parent concentration.)

Maths Numeracy was also entered as a Post-Test outcome variable in a regression on the four meeting variables, but as has been found in the path model, Maths Numeracy is a relatively weak outcome in relation to parent programme attendance. In the stringent conditions of using the whole sample, rather than the maths programme groups alone, Maths Numeracy is not predicted by either set of meeting variables.

A further and more specific test was undertaken to examine the relationships between the meeting attendances of those who participated in the parent programmes and their children's post-test attainments in reading and mathematics. This is of course the basic test of the research hypothesis that parents' attendance at programme meetings would differentially influence their children's post-test scores. Since the main focus of the study is on the effect of the intervention programmes on a disadvantaged sample, and in view of the findings reported later (in path model 2) about the nature of those advantaged parents who participated in the programmes, it was decided to carry out this final test using only those parents whose children attended one of the five schools in disadvantaged areas.

The squared correlations (equivalent to shared variance) between parent attendance at programme meetings and the children's post-test performance, more

than a year later, are given in the table below.

Table 34. Squared correlations between parent attendance and child performance

Programme Groups	Children's post-test scores on Infant Reading Test	Children's post-test scores on Maths Concepts
Attendance at reading groups (48 parents at disadvantaged schools)	0.0899*	0.0057
Attendance at maths groups (32 parents at disadvantaged schools)	0.0451	0.1059*
* $p < 0.05$ on one-tailed test of the simple correlations		

Unless the parent groups had been hand-picked on the grounds of their children's bias towards reading or mathematics, a result such as this does appear to offer fairly convincing evidence that the programmes were effective, among the disadvantaged sample, in raising the levels of reading or mathematics in the children of parents attending the reading or maths programmes respectively.

For the reading groups' scores the differences are fairly striking, with the shared variance figures showing what is virtually a zero relationship with the post-test score of maths concepts (0.006), compared to a modest shared variance with post-test reading of about 0.090 (i.e. 9 per cent variance in common). The maths groups' scores show much the same relationship with the maths concepts outcome (0.104, or over 10 per cent variance in common); this result is 2.3 times as high as the shared variance between maths attendance and post-test reading (0.045).

In the context of this latter finding it should be remembered that a number of the parents who attended the maths programme meetings said they would have preferred to attend reading programme meetings, and at least one of those parents revealed that she had borrowed reading programme materials from a friend who was attending the reading meetings. One could thus reasonably have expected some slight bias from the maths meetings variables in the direction of a positive reading prediction.

In general it is not unexpected that there should have been some small positive correlation between the level of the parents' willingness to attend programme meetings (as measured by their attendance scores) and their children's academic performance, so that it is unlikely that there would be a zero or negative correlation with performance in that subject (reading or mathematics) which was not dealt with at the programme meetings. However the sharply reduced cross-correlation with maths attendance and the near to zero cross-correlation with reading attendance do not appear to be open to any explanation other than

the differential success of the parent programmes.

Although it would be possible to explore this theme in greater depth, using a variety of other reading and mathematics outcome measures and examining correlations across sub-samples, the limitations of correlational analysis, of which mention has already been made in section 5.11, suggest that further evidence on the success or otherwise of the parent programmes should be sought in the multivariate analyses which are to follow.

6.74 ...../

#### 6.74 Correlational evidence on other variables

Correlational evidence was examined in regard to several important variables whose performance was not what might have been expected within the path regression equations, or whose performance suggested a re-examination of the correlations for other reasons.

English Picture Vocabulary Test. Evidence on the relationship between the verbally-oriented Maths Concepts measure and language came from two of the path models as well as from simple correlations. While E.P.V.T. at Nursery level correlated at 0.210 with Post-Test Maths Numeracy, the same measure of language correlated 0.464 with the Post-Test measure of Maths Concepts.

Nursery and Reception Needs. These composite variables were scored on the basis of estimates of Need for Security and Need for Esteem, as judged by the teachers in each of the six Nursery classes and in a larger number of Reception classes. Useful validation for the measures came from the correlations between the composite Needs measures derived at each class level. For the Whole Sample this was 0.455; for Disadvantaged White children alone the correlation was 0.574, and for Disadvantaged Black children alone it was 0.460 - despite the fact that the Black Needs variables predicted considerably more Post-Test attainment than did the White Needs variables. While correlations of this order between the composite variables suggest that the children's Needs levels may have altered considerably over the period between the Nursery and the Reception (Mid-Test) assessments, and/or that the measures themselves may not be too reliable, it should be remembered that each of the individual Needs measures were based on only three questions, each with simple 3-category responses. This in itself reduces the possibility of a high correlation. The subjective nature of the judgements on such questions is a further reason for not expecting too high a correlation between the composite measures of Nursery and Reception Needs.

Piagetian Tests. The slightly negative correlation between Piagetian Tests at Nursery level and the age of Nursery testing (-0.051) may to some extent be an artefact of the age distributions within the advantaged and disadvantaged samples. However even a biased age distribution could only account for a drop in the expected positive relationship. The reasonable reliability of the Piagetian Tests at Nursery age (0.78) suggests that its low correlation with age may be due to factors other than inadequate scoring of the test itself. Among the possible explanations are that children at this age (4) are on a relative plateau in relation to the measures used, or that the older four-year-olds in the sample attempted to read more into E's words than would be expected from the test protocol. The older children may have tried to draw more from the

'social logic' of the situation than from the 'formal logic', and in consequence offered more incorrect answers or manipulations than they might otherwise have done, in the belief that these were the expected responses. As with the E.P.V.T., this test also offers interesting variations in its predictive powers within different path models.

Sex group. This variable appears only occasionally as a useful predictor in the path models and has relatively low correlations with most of the tests used (with the exception of Piagetian Tests, where the Whole Sample correlation with Sex Group is 0.238).

'Non-sample school'. The correlation between total Post-Test attainment and whether or not a child had its final assessment in a non-sample school (as occurred with 23 children who had moved to 20 other schools) showed that there was no measurable relationship between departure from the sample area and attainment levels. This finding offers support for the evidence presented in section 6.10 that the differences between the attrition and surviving samples are so small that they do not impair the validity of the findings. (The attrition sample includes only those children who were totally lost to the sample; some were lost because they left the sample before the Mid-Test battery, some moved to distant towns, and a few could not be traced.)

2-weekly and 6-weekly groups. The original intention to differentiate between the groups attending 2-weekly and 6-weekly meetings was abandoned when it was realised that the 13 attenders of the 6-weekly reading programme would not be sufficient to enable a reliable path model to be developed for that group. The numbers were also too small for any conclusions as to the effectiveness of the frequency of meetings (the lower frequency meetings lasted twice as long), since there was a large difference in the numbers attending the two types of meeting. McNemar (1969) points to the possibility of Type I or Type II errors when comparing samples of radically differing sizes and differing variances, if one of the samples is particularly small. Other comments on the question of meeting frequency appear earlier in this chapter.



## 6.80 Path Analyses

The alternative approaches to analysis of the large number of variables in the study were discussed at some length in chapter 5. Some of these approaches have been pursued in the preceding sections of this chapter.

The principles underlying the major examination of the data in this final section are in essence the following:

- a. Any examination of the relationships between performance and intervention needs to rely on complex rather than simple analyses, to take into account at least part of the multiplicity of factors responsible for the outcomes under study.
- b. Interpretation of the data in this study should preferably be based on regression methods rather than on analyses of variance; the former are geared more closely to prediction and speculative interpretation, while the latter are geared more to a close examination of the probabilities of differences between groups or across categories of performance.
- c. Within the regression model, the V-ridge algorithm pioneered by Vinod (1976a, 1976b) and extensively developed in this study should be utilised rather than the ordinary least squares algorithm; theoretical and practical justification for this decision have been presented in chapter 5.
- d. Path analysis should be employed to link the results of the many separate regression equations; longitudinal models should be used rather than cross-sectional ones, with predictor variables always being antecedent in time to the dependent or outcome variables.
- e. The particular form of path analysis to be used here should be based on the concept of point prediction, developed by Wold (1975) and discussed in some detail in section 5.303. Thus separate multiple regressions should be used rather than the two-stage or three-stage least squares methods of solving simultaneous equations.
- f. The unreliability of all the variables used in the equations should be taken into account through the use of disattenuation procedures; latent variables constructed from raw variables should be seen as having their own degree of unreliability, based on a weighted average of the constituent variables.
- g. A sensitive combination of the size or meaningfulness of unique effects or predictions, and the significance of those effects, should serve as the criteria for inclusion or exclusion of variables in the path models. To cite

Namboordiri et al (1975), they point out that the main source of deviations from prediction lies in specification errors - random and non-random measurement errors, omission of relevant variables from the model, and errors due to non-linearity or non-additivity. They consider that errors such as these make significance tests meaningless, and that omitting paths which fail to reach a certain level of significance can be damaging to the model as a whole.

Details of the particular methods used in the path analyses developed here and the justification for these methods will be set out before dealing with the models themselves, as there is inevitably a close link between the methods and the possible interpretation of analyses based on those methods.

### 6.81 Application of regression techniques

The reasons underlying the application of various theoretical perspectives in the regression equations are set out below.

#### 6.811 V-ridge algorithm

While the preliminary evidence from the development of the V-ridge algorithm indicated its considerable superiority to the ordinary least squares (OLS) algorithm, the repeated cross-validation exercises on the study data have continued to show the real potential for this method. In many though not all the regression runs on the smaller samples the OLS solutions were meaningless, while V-ridge invariably offered credible solutions.

An example of these contrasting solutions appears overleaf, taken from the penultimate stage of regressing 'Total Attainment (Post-Test)' of the Working Group (N = 28) on all 12 predictor variables.

A fuller examination of this V-ridge solution showed that the performance of two of these predictors was anomalous (see the discussion of the Needs variables in Model 3, section 6.84). However this initial regression did enable a preliminary interpretation of the relationship prior to eliminating the poorest predictors. By comparison the OLS solution was totally meaningless and offered no clues as to the removal of inappropriate predictors.

With larger data sets OLS was more likely to provide at least an acceptable solution, although in many cases the OLS solutions would over-emphasise a few predictors and reduce smaller predictors to meaninglessness or provide several negative regression weights, in contrast to the generally positive solutions provided by V-ridge. In one model, for example, the OLS coefficient yielded a probability level for the Parent Programmes variable twice as high

Table 35. Comparison of regression solutions for a small sample (N=28)

Corrected mean sum of squares		1.5272
OLS residual mean square		51.0039
Variance accounted for by OLS coefficients		-.66229
V-ridge residual mean square		.3827
Variance accounted for by V-ridge coefficients		.78612
Uncorrected mult. correlation squared on the same data:		
R <sup>2</sup> for OLS	.0765	R <sup>2</sup> for V-ridge .8712
Unique contribution of each variable, as percentage of variance accounted for in Total Attainment:		
	Nursery Ability	Nursery Attain.    Receptn Ability    Parent Ac.Env.    Time in Nursery    Time in Receptn    Nursery Needs
OLS:	-641.14	-68.34    -112.06    -661.12    -473.34    -332.21    -735.47
V-rdg:	2.14	5.50    3.82    3.43    1.73    3.58    2.32
	Receptn Needs	EPVTest    Age at Post-T    Nurstat    Recstat    Shared variance contribution
OLS:	-769.59	-539.72    -3.77    -503.91    -540.98    *****
V-rdg:	7.57	0.07    3.07    2.54    0.65    40.39

as did the V-ridge coefficient. The OLS coefficient was, however, negative, suggesting a detrimental effect of the Parent Programmes on later attainment.

Gross-validations themselves invariably showed V-ridge to have a higher validity, as measured by the Index of V-ridge Effectiveness (IVE), a composite of four comparison parameters defined in chapter 5. While the V-ridge IVE was generally in the region of 2 to 8 times better than the OLS solution, in some cases the index showed performance a good deal better. The highest figure seen was 16,000. There were virtually no cases where OLS outperformed V-ridge, although on some occasions the two solutions were fairly similar.

What needs to be recognised is that without the power of V-ridge to cope with serious multicollinearity there would have been no chance of developing sensitive analyses for the relatively small number of cases in this study. It is only when data sets reach approximately 1,000 or more cases than OLS becomes competitive with V-ridge solutions.

It should also be noted that whereas the removal of non-contributing variables (such as EPVT in the above example) can lead to massive changes in the coefficients and other parameters of regressions based on OLS, such removals usually imply only modest changes in the V-ridge parameters, so that the

final reduced model in any particular case is not radically different from the model presented initially.

6.812 Stepwise, hierarchical or simultaneous entry?

Befinement of the regression equations is clearly a major element in the complex procedure of constructing path models. The serious limitations of the assumptions underlying stepwise and hierarchical regression procedures have been dealt with in sub-section 5.212. The particular deficit of the stepwise procedure is that it relies on a totally mechanistic statistical routine to determine the inclusion or exclusion of variables from the regression equation. In so far as the hierarchical method enters certain contemporaneous variables ahead of others it can also be criticised because the conceptual grounds for assuming regression priority among a set of co-predictors are hard to justify, certainly in the behavioural sciences where variables present at a specific point in time exert their influence simultaneously rather than in some imaginary sequence.

The statistically and conceptually more valid method is to enter all the predictors that are present at a certain point in time in a joint or simultaneous equation, leaving the unique variance contributions of each coefficient and the probability levels to determine whether any particular variable should be included or excluded. This is the method followed in all the equations derived for the path analyses presented here.

6.813 Minor perturbation of non-invertable matrices

With the degree of multicollinearity existing in the data it was sometimes impossible to invert certain matrices (a necessary part of the procedure for deriving regression coefficients). While the V-ridge procedure imposes a constant addition or ridge on the diagonal - the extent of this addition being determined non-stochastically within the programme - and thereby facilitates inversion of problematical matrices, a true Ordinary Least Squares solution demands an unridged matrix. In effect these contrasting procedures could mean that in a considerable minority of regressions no comparison could be made between V-ridge and OLS solutions, since no OLS solution could be obtained without a successfully inverted regressor matrix.

During the development of the V-ridge computer programme it was decided to incorporate a sub-routine which can be invoked when thought necessary to force an inversion if the original matrix cannot be inverted. This sub-routine adds a small pre-set constant to the diagonal of the regressor matrix

(customarily 0.05 was chosen for adding to the diagonal element values of 1.0). If inversion still proves impossible the programme can be instructed to repeat further constant additions (up to a set maximum) until inversion proves possible.

It should be understood that for the V-ridge procedure the effect of such a small perturbation is simply that of adding a preliminary minor ridge, obtaining the inversion and eigenvalues, and then using these to determine how much additional perturbation or ridging is needed to achieve the non-stochastic V-ridge solution. As far as the mathematical logic of this process can be judged, it does not affect the final V-ridge results (the V-ridge constant being determined on the basis of the eigenvalues of whatever matrix is presented to it). What does happen, however, is that the OLS solution is no longer an unbiased 'least squares' solution, since it is now the solution that would be obtained from a minor ridge regression. Even in these cases the V-ridge solution continues to prove superior to that of the modified OLS solution, as measured by the criterion of the IVE index.

#### 6.814 Covariance or correlation-based regression?

The V-ridge programme contains the facility for deriving both V-ridge and OLS regression solutions based on either covariance or correlation matrices. Early trials with the sample data showed that the IVE comparison was at times much better for covariance-based V-ridge solutions than for correlation-based ones. (Clearly the OLS solutions are the same whatever the matrix, whereas the V-ridge solutions differ according to the choice of matrix.)

However the correlation-based V-ridge solutions were more consistent in their improvement over OLS solutions and it was decided to adhere to correlation-based regressions throughout. It was interesting to note that although the probability statistics and size of the unique variances did vary somewhat between the covariance and correlation-based V-ridge regressions, the retention or exclusion of variables seldom differed according to the choice of matrix. The fact that some probability and unique variance differences exist are not in themselves a reflection on the reliability of the V-ridge algorithm but rather an indication that the standardisation implied in a correlation-based matrix does bring out different characteristics of the data and offer differing interpretations from those driven from covariance-based regression.

#### 6.815 Standardisation?

The issues at stake here were reconsidered at the time when the main analyses were undertaken. As explained in section 5.304, it was felt prefer-

able to rely on standardised data throughout; the arguments put forward by Hauser (1971) for using standardised data in his extensive path model - arguments cited in section 5.304 - are particularly cogent in this regard.

An additional consideration is that the Wold predictive path model which this study is following, with some minor modifications, requires standardised data for its successful employment.

#### 6.816 Squared multiple and single correlations

$R^2$ , the squared multiple correlation between all the predictors in the final regression equation and the dependent variable, is corrected throughout according to the McNemar (1969) shrinkage formula described and discussed in section 5.202.

In this regard it should be noted that while the squared correlation between a single predictor and the dependent variable may be seen as some indication of the strength of the simple relationship between that predictor and the outcome, on various occasions the total predicted variance within a multiple regression relationship is less than that of the largest of the squared correlations ( $r^2$ ) with the dependent variable. This is because a single correlational association cannot reflect a regression relationship spread over multi-dimensional space, nor does it take into account the effects of multicollinearity between predictors. The application of the McNemar formula also reduces the total 'explained variance', widening the gap with the squared single correlations even further.

#### 6.817 'Suppressor variables'

Cohen and Cohen (1975) offer an interesting discussion of the phenomenon of what they and other authors have termed 'suppressor variables'; this phenomenon arises in situations in which least squares solutions of multiple regression equations yield one or more coefficients opposite in sign to the corresponding correlations with the dependent variable. Some authors present examples in which credible explanations are offered for particular cases of this anomalous behaviour of the coefficients, arguing that one predictor variable may well have a negative effect on the outcome variable in the presence of another variable, whereas on its own the first predictor would have a positive relationships with the outcome. The reasoning is that the first predictor somehow serves as a modifier of the effects of the second predictor.

Cohen and Cohen argue that the suppression phenomenon may take different forms, depending on whether:

- a. the regression coefficient is large while the correlation is close to zero;
- b. the (standardised) regression coefficient is larger than the corresponding correlation; or
- c. the regression coefficient is opposite in sign to that of the correlation coefficient.

Darlington (1968) argues that it is wrong to try to eliminate or explain away the effects of suppressor variables, since there can be many different explanations for the phenomenon in a multivariate situation; but he too sees it as a meaningful occurrence.

The contrasting results of V-ridge and OLS solutions of regression equations in the present study suggest a completely different interpretation of this phenomenon.

In a great many equations the OLS solution includes one or more negative coefficients while the V-ridge solution is fully positive; in a small number of situations, particularly when E.P.V.T. is a co-predictor, the V-ridge solution may include a negative coefficient, opposite in sign to the corresponding correlation with the outcome. (It should be noted that the scoring of all variables was defined so as to yield positive correlations with the outcome variables.) However even such situations do not persist across the different models or samples, although if there was a genuine explanation for the phenomenon one might expect some consistency.

Of the small number of variables which at times show contradictory signs in the V-ridge regression solutions, E.P.V.T. is the most interesting. When originally grouped with other 'early attainment' variables it frequently yielded negative coefficients, and yet when retained as a separate variable and entered alongside the more powerful combinations of variables at a later stage in the path models, E.P.V.T. has often yielded a coefficient close to zero. It is noticeable that E.P.V.T. yields a coefficient that is both positive and significant for the Advantaged Sample, in virtually every equation (for that sample) in which it is entered. The implications of the finding that E.P.V.T. serves as a significant and meaningful predictor in the Advantaged Sample, but not in any of the Disadvantaged samples, is discussed later. What is important here is that the behaviour of the variable in the Disadvantaged context shows it to be of little or no importance in most situations.

While a more detailed examination of the claimed suppression phenomenon is necessary before reasonable finality can be reached as to its existence or stability, the evidence from this study, comparing the situations where OLS yields coefficients of contradictory sign while V-ridge does not, and noting also the serious inconsistency of the limited appearance of this phenomenon

within V-ridge solutions, suggests that suppression as an explanation has no real credibility. On the contrary, it appears that the apparent suppression behaviour is likely to occur when a predictor with a moderately high correlation with the dependent variable is merely duplicating the prediction of other variables; in other words it then has little or no unique variance contribution of its own.

In such a situation of high multicollinearity within the distributional space of this predictor and the multivariate space occupied by a set of other predictors, it may well occur that the overlapping relationships at times create a pseudo-suppressor effect. It seems unreasonable to propose that such a situation has any genuine conceptual meaning like that of 'suppression'; it is rather a situation of multicollinearity in which a variable with contradictory signs can be assumed to be making no real unique contribution to the model.

As a general rule in the path analyses for this study, the very small number of variables showing contradictory signs in V-ridge regressions were excluded from the particular equation in which they showed negative coefficients, although this was never done routinely but only after a close examination of the correlation matrix underlying each equation to ensure that there were reasonable grounds for the omission. It should be stressed that in every case the unique variance prediction of the variable in question was relatively small - at most only a few per cent. The fact that this prediction was so small, and often close to zero, is further evidence of the hypothesis that what one is really faced with is a particular form of multicollinearity rather than suppression in any conceptual terms.

#### 6.818 'Error' variance

The unexplained variance within any equation was determined as

$$1 - \bar{R}^2$$

where  $\bar{R}^2$  is the corrected multiple correlation squared between the final set of predictors and the dependent variable.

This is clearly a conservative estimate, since it is based on the shrunken  $R^2$  rather than on the original figure. The nature of the unexplained variance is highly problematical. Undoubtedly a considerable part of this residue will be due to unmeasured predictors which also contribute to the criterion variable, although it has not been possible to assess these within the present study. The remaining part can be seen as true error variance, representing the limitations of the test instruments and of the procedures used when employing them.



6.82      Construction of path models

In contrast with those statistical models where certainty and finality are sought in order to take cautious decisions, for example on a new medical treatment, a path model offers a sensitive and flexible statistical method for interpreting relationships and exploring possible causal links between variables. While the rules of probability and meaningfulness on which the path model is built need to be specified clearly, particularly so that others may replicate and test the work or challenge its conclusions, there is no dogmatic formula or convention which lays down exactly how a path model should be constructed or interpreted. Nor are there unwavering rules - other than those specified in advance - limiting what can or cannot be attempted within a model. The credibility of the model rests on the extent to which intuitively acceptable statistical and conceptual principles have been observed in its construction, and on the strength of the justification for ignoring certain conventions which are more generally observed by other workers in the same field.

In essence, the analyses in this study are based on a fairly wide variety of interrelated conceptual criteria to which advanced statistical models and a limited set of statistical rules are applied, rather than being based on the alternative of adhering to well recognised but rather rigid statistical procedures and criteria, within which a limited range of concepts can be tested for acceptability. The differences between the two approaches are so deep as to suggest a move from one scientific paradigm to another, from closely controlled models in which well recognised statistical procedures determine absolutely the nature of the analyses, to macro models in which wide-ranging research hypotheses serve to justify a bold employment of any legitimate techniques. Hauser (1973) was an early pioneer in this paradigm shift, presenting a highly complex model to which innovative statistical techniques were applied.

The rationale for basing the path analyses in this study on regression equations rather than on a detailed examination of 'path coefficients' (formed from partial and full correlation coefficients between the variables along a particular series of pathways) has been discussed briefly in sub-section 5.302. Given all the doubts, discussed in section 6.70 and elsewhere, about the stability of correlation coefficients, the decision to adhere to regression equations for path model construction appears justified.

There are ten path models, five of which have been built up quite independently of each other. The remaining five are termed Satellite Models; the construction of these five is closely linked to the construction of one major Base Model and they are therefore not independent, although the important diff-

erences in the contributions of different variables and latent variables within each model indicate that most of the individual sample characteristics are still evident - even within what was deliberately designed as a unifying framework. The rationale for the Satellite Models, with their advantages and limitations, are discussed in sub-section 6.825.

In that and in other parts of this section the procedures for the construction of path models are set out; these include an explanation of the meaning of the 'status variables', the limitations of the interrelationships portrayed in the model, the procedures for inclusion or exclusion of variables in each model, and the details of why and how satellite models are constructed.

#### 6.821 Basic design

The decision to adhere to the point prediction model set out by Wold (1975) has already been described. In this model the interrelationships between variables is maximised in the direction of a point prediction of a set of outcome variables, rather than maximising relationships as a whole between all the variables.

The graphic presentation of the model was given some consideration. As with other choices of method, a particular decision can influence the conceptual thinking underlying the wider issues under study. One possibility which was attempted was that of designing the model in the form of concentric circles, with the exogenous variables present at time 1 being placed in the outer circle, the exogenous and latent variables present at times 2 and 3 portrayed in the next two circles, and the final outcome variables portrayed at the centre of the circles. This would emphasise the nature of point prediction.

On the other hand it has to be recognised that the outcome variables in the present model of relationships are themselves potential input variables for a further growth in the child's performance and ability. They are not an end point in development but part of an incremental continuum. For this reason the basic graphical design is that in which the variables earliest in time are pictured on the left of the model and those latest in time on the right of the model.

For simplicity the construction of each path model is sketched on three separate sheets; two sheets define the creation of each of the latent variables for that model from the raw (standardised) variables; the third extended sheet, on which the broad model is presented, contains only the latent and status variables, together with a few raw variables not incorporated in any of the latent groupings. This makes it easier to examine the main characteristics of each model.

6.822 Detailed construction

The following steps were taken in the construction of each of the five major path models (the special procedures followed for the satellite models are discussed in sub-section 6.826):

a. Initial inclusion of variables. Decisions were made on the basis of the performance and reliability of each variable, including its conceptual importance and correlations with other variables, as to whether or not to include the variable in the initial regressions for the path model. Reasons for abandoning a few variables have been discussed in earlier sections. Various categorical variables were considered but only some of these were included, such as sex, ethnic group and programme group; apart from the sex variable, group variables were not employed as predictors but were used to sub-divide the sample for separate analyses based on the groups themselves. A variety of other variables, such as the mothers' views of nursery education, were considered but because of their uncertain relationship with parental behaviour and environment they were not included in the model.

b. Conceptual groups. The total assembly of variables was divided into conceptual groups, namely:

Nursery Ability - cognitive and meta-cognitive variables measured at the start of the study

Initial Attainment - scores on the reading, mathematical and Piagetian Tests at the start of the study

Nursery Needs - teachers' estimates of two of the Maslow needs for each sample child in the Nursery classes

Parent Academic Environment - variables derived from the Parent Interview at the start of the study

Parent Programmes - attendance scores for the different programmes

Reception Ability - cognitive and meta-cognitive variables measured on the child's entrance into Reception Class

Reception Needs - teachers' estimates of two of the Maslow needs for each sample child in the Reception classes

Post-Test Attainment - scores on the reading, mathematical and Piagetian Tests at the end of the study

For various reasons, discussed elsewhere, the English Picture Vocabulary Test was retained as a separate variable and not included in any group variable.

Time in Nursery and Time in Reception - the method by which they were created is set out in sub-section 6.63 - were also entered independently, as it is

clear that they are not conceptually part of any of the above groups. Age at Post-Test was also entered separately, although Age at Nursery Assessment and Age at Reception Assessment were incorporated in the Nursery Ability and Reception Ability groups respectively; in view of the continuing importance of the age variable it was felt appropriate to enter it again as an independent predictor in the final regression equation for the model.

c. Creation of dependent variable groups. Although the Wold model (Noonan and Wold, 1977) specifies an iterative procedure for maximising the relationship between the predictor groups and any dependent variable group - the procedure is described in sub-section 5.305 - it was decided to modify this procedure and limit the model to a specified set of weightings for the variables in each outcome group; the reasons for this are set out in the same sub-section. Thus the Post-Test Reading Attainment group variable was a combination of the three reading scores, with weights arbitrarily determined by E in relation to the assumed conceptual importance of each test in this study; the same principles applied to the composition of each of the two mathematics outcomes and the overall outcome - the latter being defined as Total Attainment (Post-Test). Details of the weighting of the individual outcome scores, to create the dependent variable groups, are presented in the next sub-section (6.83) on the specific criteria used for the models.

d. Creation of latent variables. For each path model the relevant dependent variable group was created and then used in separate multiple regressions on each individual set of predictors. Thus, for example, a particular Post-Test attainment composite variable was regressed in turn on the Nursery Ability set of variables, the Initial Attainment set, the Parent Academic Environment set, and so on. In each case the solution of the initial equation was examined and a decision made (according to the procedures described in sub-section 6.824) on the inclusion or exclusion of individual variables from the particular grouping. When this was completed the regressions were re-run on the reduced sets of variables. The process was continued until a set of raw variable predictors was obtained whose regression characteristics met the required criteria for inclusion in a grouping. The weights obtained from the final equations were used to create the latent variable for each particular grouping. The formation of all these variables is portrayed on the first two diagram sheets for each path model.

e. Creation of status variables. For reasons discussed in sub-section 6.823 and later, it was decided to create a macro latent variable representing the set of latent variables present at each stage of the path model. Thus, within the temporal sequence of each model a 'Nursery Status' latent variable

was created to represent the initial groupings of variables measured at the start of the study (in the Nursery classes and homes); the 'Reception Status' latent variable was created in turn to represent all the groupings (including the earlier Nursery groupings) based on variables scored up to the time of the assessment on entry into the Reception class. The method of creating these status variables was the same as that for the latent variables, except that the individual constituent variables were now the latent variables themselves rather than the raw variables from which the latent variables were composed. To create the Nursery Status variable the dependent or outcome variable was first regressed simultaneously on all the latent variables present at the initial Nursery stage; the equation was reduced methodically until all the remaining predictors met the criteria for inclusion. The weights obtained from that equation were then used to create a Nursery Status variable (the weights being multiplied by the respective latent variable scores). The same procedure was followed for creating the Reception Status variable. What should be noted here is that the predictors at this level included not only all the Nursery and Reception level latent variables but also the Nursery Status variable formed in the previous stage.

f. Final equation for the whole model. All the latent variables present in the model, including the Nursery and Reception Status variables and the few raw variables entered in the macro model, were used in a regression of the outcome or dependent variable on these major predictors. Again the procedure was followed of eliminating variables which did not meet the criteria for inclusion, until a final satisfactory equation was achieved. The diagram of the macro relationships within each path model is presented on the third (extended) sheet for that model.

g. Subsidiary equations for Reception level latent variables. The nature of the path model used here, with its emphasis on point prediction, raised certain problems in regard to the formation of the three Reception level latent variables, namely Parent Programmes, Reception Ability and Reception Needs. Since each of these latent variables was itself the creation of a group of raw variables related to the conceptual nature of that particular latent variable, it would have been difficult to modify the procedure so as to take into account the possible contributions of all the prior variables in the model. Thus, for example, Reception Ability is no doubt to a large extent predicted by Nursery Ability; but it may also be predicted by other variable groupings present at the Nursery level. However if these prior variables were to be included in the equation from which Reception Ability was derived, alongside all the raw cognitive variables assessed at the start of Reception, the Reception Ability variable would no longer be a variable representing only the assess-

ment of cognitive ability at the second stage. It was therefore decided to regress the three Reception level variables on to the Nursery level variables, in multiple regression equations which would be run once only, to give an approximate indication of the strength or otherwise of the Nursery level variables in predicting the three Reception level variables. These equations are portrayed on the perimeter of the third diagram sheet for each path model. It should be noted that the three equations are not refined in any way, as they are used only to indicate approximately the extent to which the Reception level variables are dependent on particular Nursery level variables.

#### 6.823 Limitations of the model

Every statistical model has its limitations; some of these are due to the conceptual limitations of the research hypotheses, while other limitations reside within the statistical procedures used. The possible conceptual limitations of the models developed in this study have been discussed at various points throughout these pages. The limitations of regression and path analysis have been referred to in chapter 5.

It is recognised that the present path model is not the fullest that might be constructed from the data, but given the available resources of time and effort and the fact that a number of new techniques are being pioneered in this study, the existing limitation does not seem unreasonable.

The creation of the Status variables makes up to some extent for the decision not to incorporate the Nursery latent variables within the three Reception level variables, since these Status variables do group the power of the prior set of variables into a new latent variable, enabling in certain cases the elimination from later equations of some of the earlier predictors.

#### 6.824 Procedures for inclusion and exclusion

Sub-section 6.822 has described the steps taken to construct the path models. One of the most crucial elements in the construction is the procedure for inclusion or exclusion of variables. The criteria used for these decisions include the unique variance predicted by a variable in competition with other predictors in the model, and the probability level of the regression coefficient of the variable in question.

An important finding from the V-ridge regression algorithm is that it provides an intuitively acceptable set of unique variance statistics for all the variables in an equation. Thus a decision on variable inclusion or exclu-

sion can be based not only on the customary statistic of the probability of each coefficient (or the probability of the variance added at each step, in stepwise regression), but also on the size of the unique variance contributed by each variable. A typical example illustrates the difference between V-ridge and OLS regression in the interpretation of unique variance. The dependent variable in this case is a combination of all the post-test reading and mathematics measures.

Table 36. Comparison of probability and unique variances in a path regression

Sample: Total Disadvantaged (N=129) Predicting: Total Attainment (Post)									
Regression on all the variables present at the second level in the model									
Latent/Raw Variables	Nursy. Ablty.	Initl. Attan.	Recep. Ablty.	Parent Acadm.	Time Nursy.	Nursy. Needs	Recep. Needs	E.P.V. Test	Nursy. Status
OLS coefft.	-.055	.357	.317	-.009	.194	-.104	.009	-.106	.397
V-rdg coefft.	.111	.230	.239	.026	.152	-.028	.022	-.010	.221
OLS problty.	.682	.000	.001	.556	.000	.946	.442	.913	.003
V-r problty.	.000	.000	.000	.225	.000	.790	.262	.614	.000
	Shared Varce.								
OLS %	61.11	0.05	2.69	1.97	0.00	2.39	0.54	0.00	0.38
V-r %	38.83	2.04	7.47	8.42	0.09	3.12	0.10	0.06	9.96

The probability data suggest that the two algorithms differ considerably in the interpretation of the importance and even of the sign of initial (Nursery) Ability but that there is fair agreement between them on the other four principal predictors. The unique variance contributions determined by each algorithm are as set out in the last two lines of the table.

The most noticeable contrast between these latter figures is also in the contribution of Nursery Ability, which is close to zero in the OLS solution - something which could be expected from an examination of the probability data.

Experience in numerous regressions has shown that whereas the elimination of weak or zero predictors from an OLS equation requires cautious removal of one or at most two variables at a time (a caution emphasised by virtually every author on this subject), in view of the considerable changes in some of the parameters as variables are removed, the removal of three or even four variables at a time causes no major changes in the V-ridge parameters, suggesting the considerable stability of the statistics derived from the latter algorithm.

The final equation in the regression initiated above was as set out in the table overleaf.

Table 37. Probability and unique variances in reduced path regression

Sample: Total Disadvantaged ( N=129 )		Predicting: Total Attainment (Post)				
Regression on reduced set of the five most powerful predictors in first equation						
Latent variables		Nursery Ability	Initial Attainment	Reception Ability	Time in Nursery	Nursery Status
OLS coefficients		-.833	.181	.627	.325	.862
V-ridge coefficients		.081	.251	.264	.164	.233
OLS probability		.9997	.2803	.0008	.0000	.0703
V-ridge probability		.0013	.0000	.0000	.0000	.0000
	Shared Variance					
OLS variance %	56.40	3.20	0.09	2.76	5.83	0.58
V-r variance %	4.78	1.53	11.05	12.27	3.59	38.04

An examination of the startling contrasts between these two solutions indicates the considerable power of V-ridge in establishing which are the important regression relationships, freed of much of the inherent error and instability in the OLS algorithm. There are a number of points of difference.

i. Nursery Ability is treated by OLS as a highly significant negative contributor to outcome variance. This would be interpreted by some authors as an example of the suppression phenomenon, although the conceptual justification for regarding Nursery Ability as a suppressor of the other predictors of attainment is not at all clear. In contrast, V-ridge presents Nursery Ability as only a moderate (though positive) contributor; this result could be expected in the presence of Reception Ability, based on the same set of tests administered at a later point in the study - with the latter variable being closer in time to the post-test phase.

ii. Initial Attainment makes virtually no unique contribution to variance in the OLS model, whereas it is a highly important contributor in the V-ridge model. All the V-ridge models in this study suggest that the Initial Attainment measures of the disadvantaged children are a persistent and powerful contributor to final attainment. Even for the advantaged children, where parent environment contributes highly, Initial Attainment still makes a moderate and persisting contribution. Intuitively it could be expected that Initial Attainment should contribute to the final variance of Post-Test Attainment, beyond what ability (cognitive and meta-cognitive skills) can contribute, and yet the OLS solution presents this variable as virtually meaningless. If a further rigorous reduction of the present OLS equation were to be applied (on the basis of the probability statistics), Initial Attainment would be the first variable to be eliminated.



iii. The contrasting variance contributions of Nursery Status raise further doubts about the acceptability of the OLS solution. Nursery Status is a weighted combination of the Nursery level variables, based on weights derived from the V-ridge solution to an earlier regression. As such this variable could be expected to contribute considerably to final attainment in the V-ridge solution. On the other hand the unique OLS variance predicted from the same variable is only half a per cent, despite the fact that the OLS coefficient for the status variable is the largest of all the coefficients in the OLS solution.

iv. The alterations in the two solutions (OLS and V-ridge) with the reduction in the number of variables also points to the greater stability of the V-ridge solution. The biggest change in the latter is that the shared variance alters from 39 per cent in the full equation to under 5 per cent in the reduced equation. While the individual variance statistics alter somewhat (particularly that for Nursery Status, which increases from 10 to 38 per cent), on the whole they do not present a radically differing view of the model across the extended and reduced solutions. In the equivalent OLS reduction the shared variance remains much the same, but some of the other parameters alter from meaningless to meaningful levels, or vice versa. A further consideration is that V-ridge tends to share out a lot more of the variance, while OLS tends to attribute nearly all the prediction to shared variance, despite the differing nature of the regressor variables.

What is important about this example, which has been taken at random from the many hundreds of similar regression contrasts in the study, is that it indicates how seriously deficient are the OLS regressions even on models as sizable as 129 cases. It also suggests that the conclusions from a great many earlier research studies, about the significance of different contributors to school attainment, may need some amendment. Even if the validity of this and other contrasts between the two algorithms could be challenged on other grounds, the added knowledge that V-ridge almost invariably presents a set of coefficients which perform more efficiently in the cross-validation test suggest that the latter is overall the more reliable algorithm.

In the two regression contrasts whose statistics were set out above, the Index of V-ridge Effectiveness (summarising the four contrasts of performance between OLS and V-ridge coefficients in a cross-validation exercise) could not be derived in the first regression because the OLS stability parameter was negative. In the second reduced regression the IVE measure showed V-ridge to have an overall superiority factor of 20.82 over the OLS coefficients.

The example set out on the previous pages illustrates the reduction of a regression equation to the level where each of the remaining predictors meets certain criteria of probability and meaningfulness (unique variance). The

actual criteria used in each path model will be set out in section 6.83, prior to the description of the 10 models themselves in section 6.84.

In essence, the V-ridge criteria of the probability of each regression coefficient and the unique variance contributed by that coefficient are used to determine the retention or exclusion of variables from the equation. Both criteria are applied jointly, so that a variable coefficient needs a satisfactorily low probability as well as a moderately sized unique variance in order to be retained in the equation. The statistics listed on the diagrams for the various path models (as well as the fuller tables in Appendix D7) show that in a number of cases variables were excluded because of failure to meet one of these two criteria, even though the other criterion was met satisfactorily.

Decisions as to the acceptable levels of probability and meaningfulness were made on a predetermined basis, in relation to the size of the sample. Thus for smaller samples the criteria were less rigid than for larger samples.

While decisions as to the meaningfulness of a unique variance are relatively easy to justify - for example, 0.3 or 0.5 per cent minimum variance could be seen as a reasonable criterion - the decisions about probability may be more contentious, since the levels of the latter were often set above the conventional level of 0.05.

This issue has already been discussed at length in section 5.16, on the significance concept, and elsewhere. As Namboodiri et al (1975) point out: "If (specification) errors predominate, it seems to make little sense to rely heavily on significance tests.... Thus our focus is on flexibility and the importance of theory building as well as theory testing. ....it seems wise to rely on multiple criteria of evaluation rather than on a single overall test or measure of goodness of fit.... If one does find it necessary to utilise a general criterion for use.... it seems sensible to adopt some departure from zero (or whatever value has been predicted) rather than relying on a test of significance which puts the premium on sampling errors alone." (p.459)

Decisions about probability levels in this study were not lightly taken. Following early experience with the regression equations and the nature of the data, probability levels were set in advance of carrying out the path analyses themselves. These levels are clearly specified in section 6.83. Only one-tailed probability statistics are used, since the design of each measure is such that all contributing variables are expected to make a positive contribution to outcome variance. Variables which fail to predict positively are rejected from the model. The relatively few instances where this occurs are discussed when the models are analysed.

The probabilities themselves will not be described in terms such as

'significant' or 'highly significant' (conventionally 0.05 and 0.001), but will be presented simply as p statistics (e.g. p .062, p .000). Figures such as p .000 indicate a probability level below 0.0005. Statistics on unique variance contributions will be presented as percentages, with a zero before the decimal point when the percentage is below 1.0, followed by the percentage sign. Correlations are likewise presented with a zero before the decimal.

One final point is that the regressions frequently show that a variable with a relatively modest squared correlation with the outcome or dependent variable may have a higher and more meaningful unique variance contribution than some other variables which have relatively high squared correlations with the outcome. Such examples may simply indicate that the siting of the first variable in the total variable space may be more independent of other co-predictors than are the remaining variables, and because of this the first variable may merit retention in the model. This point is made here to emphasise that while the initial correlation matrix was studied in order to help decide on the inclusion or exclusion of variables from the model as a whole, the correlational criteria for inclusion were not stringent and variables were retained if it was thought that they might have any chance of contributing to prediction in the regression equations.

#### 6.825 Satellite Models

One of the main goals of this study has been to examine how the prediction patterns of each Post-Test attainment differ between sub-samples, and the extent to which the Parent Programmes variable contributes to that early school attainment.

One apparent difficulty with this task is that the path models have to be built up sequentially, starting from the creation of the set of 'Nursery level' latent variables, regressing the outcome or dependent variable on all these latent variables simultaneously, then proceeding similarly to create and assess the 'Reception level' set, and finally completing the model by regressing all the latent variables together. If each path model is constructed separately the degree of comparison that is possible between models becomes increasingly complex as the time sequence develops.

This causes no real problem when comparing, for example, disparate programme samples or widely differing social samples such as the advantaged and disadvantaged groups, since it is anticipated that there may be wide differences in the basic character of each of these path models. The interpretation of the differences affords much of the material for discussion of the findings of the study.

On the other hand, when comparing disadvantaged girls with disadvan-

tagged boys, or disadvantaged Black children with disadvantaged White children, it seemed more appropriate to keep each of the group models in close parallel so that the differing predictive weights of the variables at any one level could be compared directly, rather than permitting each model to develop in what might be different directions over each succeeding level of complexity.

Not only was this parallel development a useful simplification of what is a moderately advanced set of procedures, but it also served to limit what was already a very large number of regressions and other forms of data manipulation.

Thus, while a totally separate model for every group would have been the ideal, both practicality and the aim of facilitating comparisons suggested that for linked groups a base sample should be used which included all the groups, treating the path model for each subsidiary group as a 'satellite model' linked by certain parameters to the Base Model.

It was decided that the Total Disadvantaged sample (N=129) should serve as the base sample for comparisons between the satellite samples of girls and boys or of Black and White children (Black in this context referring only to children of West Indian or African descent). It was also decided to include the disadvantaged Working Group as a fifth satellite sample rather than treat it separately. The latter could not however be paired for comparison with the 'Non-Working Group' within the Total Disadvantaged sample, since the 'Non-Working Group' included a variety of disparate sub-samples such as two different Parent Programme groups and a non-attenders group. By treating Working Group simply as one of the satellite samples, it could still be compared with the Total Disadvantaged sample of which it formed a part. (As explained earlier, in view of the smallness of the non-attenders sample - 21 - it was felt inappropriate, statistically, to develop a path model for that group alone.)

Total Post-Test Attainment was used as the outcome variable for the Base Model (and consequently for its five satellite models). The uniformity in the construction of the latent and status variables - these being entirely based on the regression coefficients of the Total Disadvantaged sample - presented the possibility for fairly clearcut comparisons between each satellite group's regression parameters at all three levels in the models (Nursery, Reception and Post-Test).

It is particularly important to note that the preliminary diagrams illustrating the hypothetical creation of latent variables for each satellite model in fact illustrate only the regression coefficients for the satellite sample. The creation of all the latent variables (for the Base and satellite models) employs the Base Model coefficients and applies these uniformly to all the disadvantaged cases, including those in the satellite samples. The exact procedure will be described shortly.

Comparisons between the latent variable diagrams for the Base Model and those for the satellite models show that in general the satellite procedure does not seriously distort the findings; if anything, it facilitates direct comparisons at a relatively small price in sacrificing the uniqueness of independent or separately created models. Another strength of the satellite model principle is that since the Base Model is constructed from a sample of 129 there is greater stability in the results and in the nature of the latent variables created from that relatively large sample.

Further discussion of the use and the merits or demerits of the satellite models will appear in the examination of the models themselves.

A brief explanation follows of the principles applied to the construction of the Base and satellite models; reference is also made to some possible anomalies.

i. Standardisation of data. The data for all 129 Disadvantaged children were standardised over that sample. However it should be noted that since the particular V-ridge regression procedure used here was based on the correlation matrix, each group's sample data were again separately standardised prior to determination of the group regression coefficients.

ii. Construction of latent variables. In preparing each latent variable for the Base Model the outcome variable was also regressed separately on to each group's set of raw predictor variables, as portrayed in the diagrams on the creation of the latent variables. However only coefficients from the Total Disadvantaged sample were used when creating the latent variables for all the samples (Base and satellite). Thus the composition of the latent variables is not unique to any sub-sample but is common to them all.

iii. Omission of meaningful variables when constructing latent variables for a satellite model. The procedure outlined above holds the possibility that in certain cases a variable which is a meaningful contributor within a satellite sub-sample may not appear in the Base Model if it is not a meaningful predictor within the latter model. This possibility will be closely examined when discussing each of the satellite models. (An examination of all the satellite models shows that meaningful contributors were omitted in only a very small number of cases. These omissions are highlighted on the main path diagrams. In no case was the omission so serious as to challenge the basic conclusions from any model.) A further consideration is that the level of inclusion of variables in the Base Model ( $p$  not greater than 0.15 and unique contribution not less than 0.5 per cent) is relatively generous.

iv. Inclusion of non-meaningful variables in constructing latent variables for a satellite model. The reverse process from that outlined above may well

occur and has in fact been noted in a number of satellite models. Variables which within a sub-sample do not make a meaningful contribution to the outcome variable have nevertheless been included in the construction of latent variables for all models because of the demands of the Base Model, where such variables do play a meaningful part. Wrongful inclusion of a meaningless or insufficiently meaningful variable may slightly reduce the power of the latent variable in the context of a particular satellite model. It is however less likely to distort results than when a meaningful variable is omitted as a result of applying the satellite principle.

#### 6.826 Reliability of predictions

At various points in this and the preceding chapters the possible limitations of the study, such as the measurement of the variables, the conceptual principles on which the path models have been constructed and in particular the statistical techniques used - some of them totally new - have been discussed. Among the remaining questions that might be posed is the problem of whether the nature of the techniques themselves could possibly be so unreliable as to inflate the results to a level well above reality.

Fortunately the nature of variance prediction - the use of a regression equation to divide the variance of the outcome variable among a set of potential predictors - is such that it is virtually impossible to exaggerate the total prediction as measured in variance explained, provided the necessary correction formula is used. Clearly there is still room for argument as to the accuracy of the estimate of the relative proportions contributed by different predictors.

It can be noted moreover that although OLS and V-ridge solutions differ very considerably in detail, the total predicted variance does not differ much between the two algorithms. With some of the models the outcome variance predicted is as high as 75 per cent, indicating that a moderately useful set of initial attainment tests, cognitive scales and parent environment measures have been assembled for this study.

## 6.83 Specific criteria

In this section the specific criteria used in the ten path models are set out in detail, together with the justification for the decisions taken.

### 6.831 Construction of the outcome variables

The initial intention was to have three separate outcomes, one comprising all the post-test attainment variables, one consisting of the post-test reading variables and the other of the post-test mathematics variables.

It was hoped that the iterative path model construction, based on the Wold (1975) and Noonan and Wold (1977) point predictive model, would help determine the most favourable combination of the various reading and mathematics variables - relating these in the highest possible level of shared variance with the various sets of outcome variables. As explained previously in sub-section 5.305, the attempt to iterate this model to a point of stability (i.e. where the weighting of the component outcome variables would not alter beyond a minimum figure with each succeeding round of the Wold algorithm) resulted in an unacceptably high weighting of the reading variables compared to the mathematics outcome variables.

It was therefore decided to fix the weights of the component variables in the overall Total Attainment (Post-Test) variable. These weights were based on the proportions set for the subsidiary reading and mathematics Post-Test outcomes. The weights for Total Attainment were as follows:

Standardised Infant Reading Test	0.2
" Southgate Reading Test	0.2
" Daniels and Diack Sentence Reading Test	0.1
" Maths Numeracy (modified WPPSI)	0.166667
" Boehm Concepts (modified)	0.166667
" Piagetian Tests	0.166667

The experience obtained with the iteration attempt on the Total Attainment variable suggested that a similar maximisation of relationships with the post-test reading variables would not necessarily agree with E's conception of what he and others might see as evidence of reading attainment at the end of the Reception year; for example, the iteration process might well have given the newly devised Infant Reading Test a very heavy weighting compared to the other two reading tests, although it is the latter which are known and accepted as reasonable measures of infant reading skills.

It was therefore decided to base the Reading Attainment (Post-Test) variable on the following weightings:

Standardised Infant Reading Test	0.4
" Southgate Reading Test	0.4
" Daniels and Diack Sentence Reading Test	0.2

The Daniels and Diack was given only half the weighting of the other tests for the reason that many children were still scoring close to zero or zero at the end of the Reception year (the Daniels and Diack test being based on the ability to read very simple sentences rather than words); there were however sufficient positive scores - approximately half the children - to justify its inclusion as a minor measure within the composite.

The decision to separate the mathematics scores into two separate outcomes was made for three reasons. Firstly, it was found that some children tended to have differing levels of performance on the modified Boehm Concepts test and on the modified WPPSI Arithmetic Test, although both formed part of the original measure of mathematics attainment, pre and post-test. The whole sample correlation between these two measures is 0.49 (shared variance of less than 25 per cent), pointing to a considerable degree of difference in scores. Certain children appeared to have a grasp of the spatial concepts involved in the Boehm measure, but not of the numeracy demanded by the WPPSI measure; the reverse applied to some other children. There were also children who scored well on the Piagetian concepts but had little numerical awareness.

Secondly, it was considered that a combination of the Piagetian tests and the Boehm Concepts test would give a satisfactory measure of a child's current awareness of what might be termed 'mathematical concepts', since it is reasonable to regard both these measures as precursors of numerical and other mathematical skills. Whether Piagetian Tests can rightly be considered as measures of attainment is naturally a debatable issue; it may well be argued that they fall into the category of abilities which are not easily taught but develop by a combination of maturational and environmental effects. On the other hand the considerable literature on the 'training' of Piagetian skills (Flavell, 1963, and Brainerd, 1974, for example) suggests that it may not be unreasonable to treat them here as partial indicators of the child's level of mathematical concept attainment and therefore open to further development as a result of the mathematics programme given to the parents of Nursery children. It can be noted additionally that the Pre-Test Piagetian measure correlates more highly with the Infant Reading Test, the Mathematics Tests (Arithmetic and Boehm) and E.P.V.T. than with any of the set of 'cognitive' pre-tests. In the circumstances it does not seem unreasonable to use it as a measure of attainment.

The third reason for separating the components of mathematical attainment



was that attendance at mathematics programmes appeared to have a much better relationship with post-test Boehm and Piagetian concepts than with post-test Arithmetic. While inclusion of concepts and numeracy within the same maths outcome variable might have increased the chance of showing that the mathematics programme had an overall though reduced effect on early maths skills, it seemed more appropriate to separate out the components of early mathematical attainment into the two conceptually differing measures and built separate path models focusing on numeracy and concepts respectively, with Piagetian Tests added to Boehm Concepts to give an overall Concepts measure.

The weights for the two measures involved in Post-Test Mathematics Concepts were as follows:

Standardised Boehm Concepts (modified)	0.5
" Piagetian Tests	0.5

This left the final measure of Post-Test Mathematics Numeracy to be simply:

Standardised WPPSI Arithmetic (modified)

It should be noted that the modified Boehm Concepts Tests, described earlier in this study, consists of a selection of the most mathematically appropriate items from the measure, with the addition of a few new pictorial items to examine awareness of the meaning of terms such as circle, square and triangle. The modification of the WPPSI Arithmetic Test, also referred to earlier, consists of the addition of five items to extend the tests's ceiling.

#### 6.832 Limits set for probability and unique variance

The question of how much importance should be attached to significance concepts has been discussed in some detail in section 5.16 and earlier in the present chapter. In brief, the arguments previously cited - supported by various writers - are that probability levels cannot be used to 'prove' hypotheses and should merely serve a cautionary function; thus, for example, findings with too high a level of chance probability would need to be disregarded. On the other hand, as Namboordiri et al (1975) emphasise, the conventional level of significance is not necessarily the criterion by which particular paths of relationships in a model should be accepted or rejected.

While the probability level of 0.05 has been breached in several of the models in this study, the total number of occasions when use is made of path (regression) coefficients with a probability higher than the conventional level of significance is relatively limited. The particular conditions applying to

the construction of the five satellite models (which are dependent on the probability criteria of the Base Model) have already been referred to; when individual satellite models are discussed these conditions will be reconsidered. The present discussion refers only to the five full models in which probability levels are used as one of the two main criteria for acceptance or rejection of a path.

The probability and unique variance criteria set for these five models, and the occasions on which the level of  $p$  0.05 has been deliberately breached, are set out in the accompanying table.

Table 38. Probability and unique variance limits for path models

Model	Creation of Latent Variables				Construction of Path Models			
	Prob. limit $p \leq$	Unique Var. % minim.	Number of paths	No. of paths GT. $p$ .05	Prob. limit $p \leq$	Unique Var. % minim.	Number of paths	No. of paths GT. $p$ .05
Total Disad. (129) predicting Total Attainment	.15	0.5	27	3	.05	0.7**	15	0
Advantaged (27) predicting Total Attainment	.30	0.3*	20	8	.20	0.7	19	2
Dis.Read.Gps. (48) predicting Post Reading	.25	0.4	25	11	.05	0.7	14	0
Dis.Maths Gps. (32) predicting Post Numeracy	.30	0.3*	23	13	.20	0.7	13	3
Dis.Maths Gps. (32) predicting Post Concepts	.30	0.3	26	8	.20	0.7	16	1
* None of these paths were below 0.5 per cent								
** One value of 0.4 per cent included								

There are two issues which can be considered here. In the case of latent variable creation the maximum probability level was set fairly high, so as to ensure that raw variables which might have a useful contribution (usually with a unique variance contribution of at least 0.5 per cent, although dropping lower than this on a very few occasions for two of the models) would be included in the creation of the latent variables themselves. In the interests of the completeness of each model these exceptions could be justified, since the ultimate test of the model would be the strength of the latent variables themselves as predictors within the total path models.

In the construction of the main path models the breaches of the  $p$  .05 level require more detailed consideration, as these models are the basis for

the major interpretations of relationships within the samples.

The first of the five full models (Total Disadvantaged sample, predicting Total Attainment) appears to be quite satisfactory, with all paths lower than  $p .05$ . Only two of the 15 paths in this model have a unique variance prediction below 1.0 per cent, one at 0.4 and the other at 0.9 per cent.

The second model (Advantaged Sample, predicting Total Attainment) has two out of 19 paths at a level above  $p .05$ . Both these paths are from the Parent Programmes latent variable, with  $p .08$  and a unique variance of 1.1 per cent at Reception level, and  $p .10$  and a unique variance of 0.8 per cent at Post-Test level. Given that the Advantaged Sample consists of both reading and mathematics programme groups, while the Parent Programme latent variable is made up only of Reading Programme attendance data - for reasons to be explained in the discussion of the model - the exception made here in regard to the probability level seems reasonable.

The next full model (Disadvantaged Reading Groups, predicting Reading Attainment) appears to be as satisfactory as the first model, with all paths lower than  $p .05$  and all unique variances above 1.0 per cent. This is an interesting model, given that it is also one of the main tests of the research hypotheses in this study.

The following model (Disadvantaged Maths Groups, predicting Maths Numeracy Attainment) has three out of 13 paths above the  $p .05$  level, although in no case is the unique variance below 0.7 per cent. The three paths for which the level of  $p .05$  was breached were:

- a. Parent Academic Environment, predicting a unique 0.8 per cent with a probability of .193 at the Nursery level; this L.V. failed to predict at the two subsequent levels (Reception and Post-Test).
- b. Parent Programmes, predicting a unique 0.9 per cent with a probability of  $p .158$  at the Reception level (the first point at which this latent variable is entered); again this variable failed to predict at the subsequent (Post-Test) stage.
- c. Nursery Needs, predicting a unique 0.7 per cent with a probability of .215 at the Nursery level; this latent variable also failed to predict at the two subsequent levels.

The final model in this table (Disadvantaged Maths Groups, predicting Maths Concepts Attainment) has one out of 16 paths above the  $p .05$  level, although in no case is the unique variance below 0.8 per cent. This particular path is the final entry of Nursery Ability (at Post-Test level), where it adds 0.8 per cent to the total predicted variance, with a probability of .144. Given the size of this contribution where six other variables are also making contri-

tions, the inclusion of the path appears reasonably justified.

While the inclusion of these six paths, out of a total of 77 paths in the five full models, may be questioned, it can be pointed out that the inclusion of such paths has not fundamentally altered the pattern of relationships in any model nor has it added any larger amount to the overall predictions; on the contrary, the poorer the predictor the less likely it is that its contribution to the outcome variable can remain in the model at each successive stage. The inclusion of the six relatively weak paths does nevertheless add insight and their exclusion would be a loss to the completeness of the relevant models.

#### 6.833 Other criteria

The data used in this study are set out in full in Appendix D2.

The derivation of the reliability figures for the disattenuation process has been described in section 6.51, and the reliabilities themselves are listed in Appendix D6. As explained earlier, the reliabilities of latent variables (including the constructed composite variables for post-test attainment) are determined on the basis of the reliabilities and weights of each of the constituent variables used in the construction of a particular latent variable.

While the nomological redundancy index is derived for each regression question, the index itself has not served as the criterion for the acceptance or rejection of a particular equation. However the indices for the three principal equations in each path model are printed in the bottom right hand corner of the model, to show the strength or weakness of these equations. Clearly such indices, in combination, are a useful quantitative measure of the nomological validity of the path model as a whole.

One finding which may seem perplexing is that the total variance predicted at the Reception Status level in some of the models is slightly higher or the same as the prediction at the Post-Test level - where one or two additional variables may have been added (Age at Post-Test and Time in Reception). This is essentially a statistical artefact, linked to the fact that while the additional contribution of these two variables may be quite small (if they do predict), the correction formula for adjusting the total variance explained imposes a stringent reduction on this latter parameter, taking account of both the number of cases and the number of variables entered into a particular equation. In addition, the nature of the process used to accumulate variable contributions at each level, forming a conglomerate 'status' variable as input to the next stage, is such as to add a new contributor to the regression at each subsequent level; this necessitates in any case a higher correction, which may also marginally lower the total variance prediction.

The reductions in variance explained between the second and third stages, as a result of the stringent correction procedure, are very minor, however, and do not affect the interpretation of the models in any way. At the same time this does emphasise how rigorous are the statistical checks built into the analysis as a whole, with a bias against undue inflation of the results.

# 6.84 Path Models 1 to 10

The path models developed from this study are set out and discussed on the following pages:

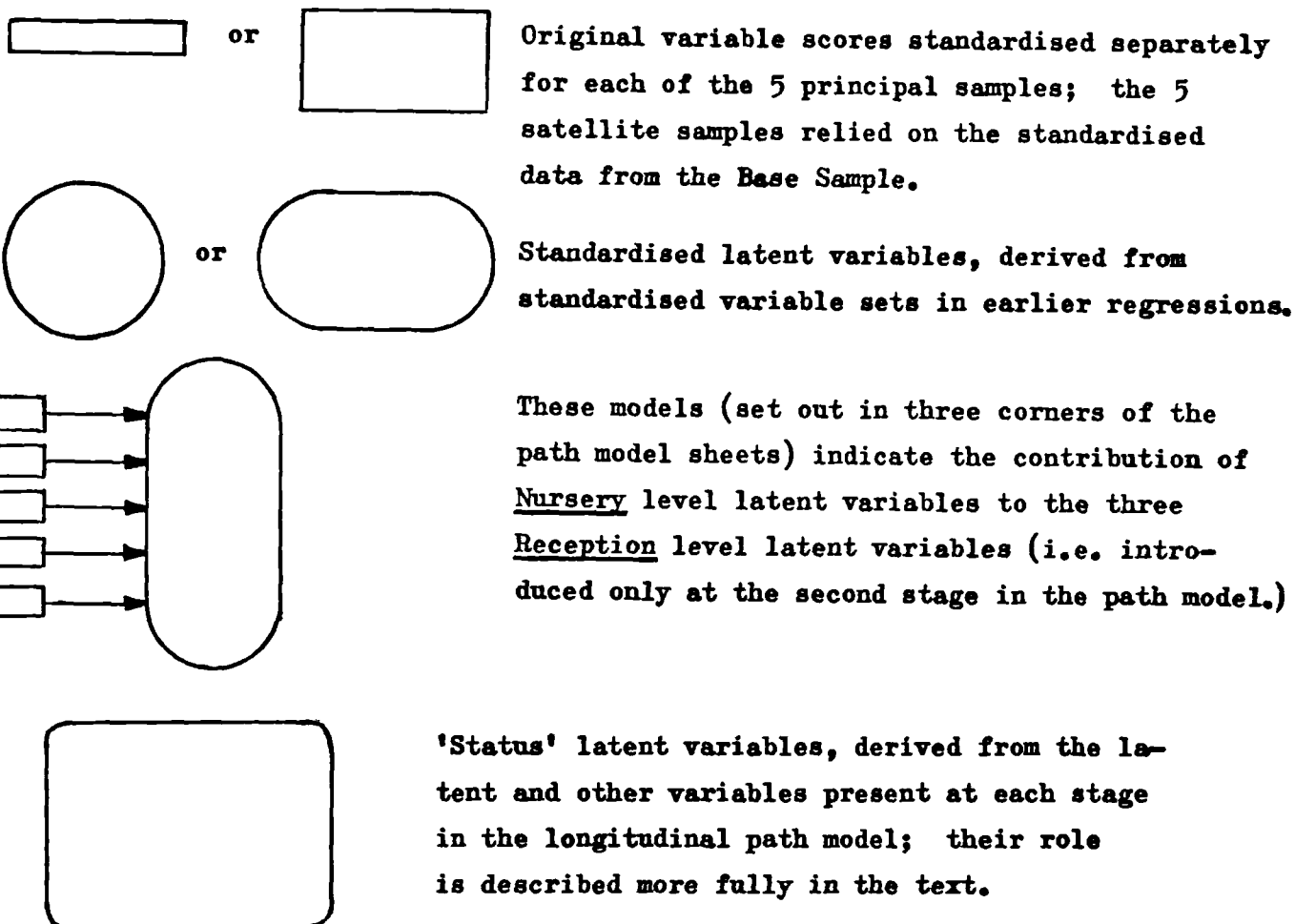
## Model

1	Total Disadvantaged (129) (Base Model for Satellites)	Total Attainment (Post-Test)	516-526
2	Advantaged Sample (27)	Total Attainment (Post-Test)	527-538
3	Disadv. Working Group (28) (Satellite Model)	Total Attainment (Post-Test)	539-548
4	Disadv. Reading Groups (48)	Reading Attaint. (Post-Test)	549-559
5	Disadv. Maths Groups (32)	Maths Numeracy (Post-Test)	560-575
6	Disadv. Maths Groups (32)	Maths Concepts (Post-Test)	
7	Disadvantaged Girls (71) (Satellite Model)	Total Attainment (Post-Test)	576-588
8	Disadvantaged Boys (58) (Satellite Model)	Total Attainment (Post-Test)	
9	Disadv. Black Children (44) (Satellite Model)	Total Attainment (Post-Test)	589-604
10	Disadv. White Children (62) (Satellite Model)	Total Attainment (Post-Test)	

The total sample available for analysis at the end of the Post-Test phase was 159.

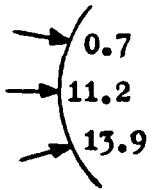
Statistical algorithms: Regressions were based on V-ridge regression rather than on the Ordinary Least Squares method. The path model was based on point prediction. The reasons for these choices are set out in full in Chapter 5.

# Legend



- $\beta$  Standard regression coefficients, presented as  $\beta .113$ , for example, along the line linking two variables.
- $p$  The figure presented (e.g.  $p .003$ ) gives the probability level of the particular regression coefficient or other parameter.
- $p .000$  This indicates a probability level below .0005; probability levels between .0005 and .001 are presented as .001.
- $r^2$  Where a variable is excluded from a regression equation because its coefficient has too high a probability figure or its unique variance is too small, the squared correlation between the variable and the criterion or dependent variable is presented (e.g.  $r^2 .015$ ).
- $u$  Where a variable is excluded (as in the previous example), the unique variance which it contributed in the original equation, prior to any exclusions, is presented (e.g.  $u 0.1$ ). Note that all unique variances are presented as percentages, and for this reason the format 0.\_ is used; for the three symbols previously described the first zero (before the decimal point) is omitted.

Legend (continued)



Figures displayed inside a latent variable perimeter indicate the unique predictive variances of each contributing variable, written as percentages of the total variance of the criterion variable. These figures are the same as the 'u' parameters described above, and are also written with a zero or other integer before the decimal point.

**Shared Var.** That part of the predictive variance shared between the individual contributors to a particular latent variable. Again this is a percentage of the total variance of the criterion variable (not a percentage of the explained part of that variance).

**Unexplained Var.** That percentage of the total variance of the criterion which has not been accounted for in the regression of the criterion variable on to the particular set of contributor variables.

Note: Most parameters have been reduced to three significant digits for presentation in the path diagrams. The data and derivations have of course been used in full within the computer programmes. Thus derivations based on the rounded numbers appearing here may at times differ slightly from the derivations produced by the computer. Figures such as 0.000 may be a rounding from 0.00047... Seemingly identical parameters may differ in subsequent digits.

\* The three starred latent variables (introduced at the second stage of each path model) have subsidiary 'check' equations whose parameters have been set out on the periphery of the main diagram. These checks are intended to establish the extent to which the starred variables may have been influenced by the first stage latent variables.

+ Slight over-prediction indicated by small negative shared variance. Unique contributions have been corrected for this.



++ Serious over-prediction, due to presence in original equation of variable(s) predicting in the opposite direction to that of their correlation with the dependent variable. As this is simply a subsidiary 'check' equation it was not necessary to derive the final equation, from which the poor predictors would normally have been removed.

+++ For reasons explained in the text, the Parent Programme variables did



Legend (continued)

not form part of this model.

 or  Either of these symbols (on a satellite model diagram) indicates that the predictor variable to which it refers is a meaningful predictor of the outcome variance in the satellite sample, even though in the Base Model the predictor in question does not make a meaningful contribution at this level. Because it is a satellite model the status variable does not incorporate any weighting from the particular variable. However the importance of that predictor within the satellite model needs to be recognised, and this is discussed under the relevant model.

Model 1

Total Disadvantaged Sample ( 129 )

predicting

Total Attainment (Post-Test)

**Sample:** All the children in the Nursery classes of five randomly selected schools in disadvantaged areas of an inner urban education authority. The only children excluded from this sample were those few whose parents could not be interviewed, as described in section 6.22. Data on one child whose parents said (after being interviewed) that they would not be interested in attending programme meetings, were excluded from the analyses.

The 129 parents in this sample include:

Parents who attended reading group meetings (48)

Parents who attended mathematics group meetings (32)

Parents who agreed to attend group meetings but failed to do so (21)

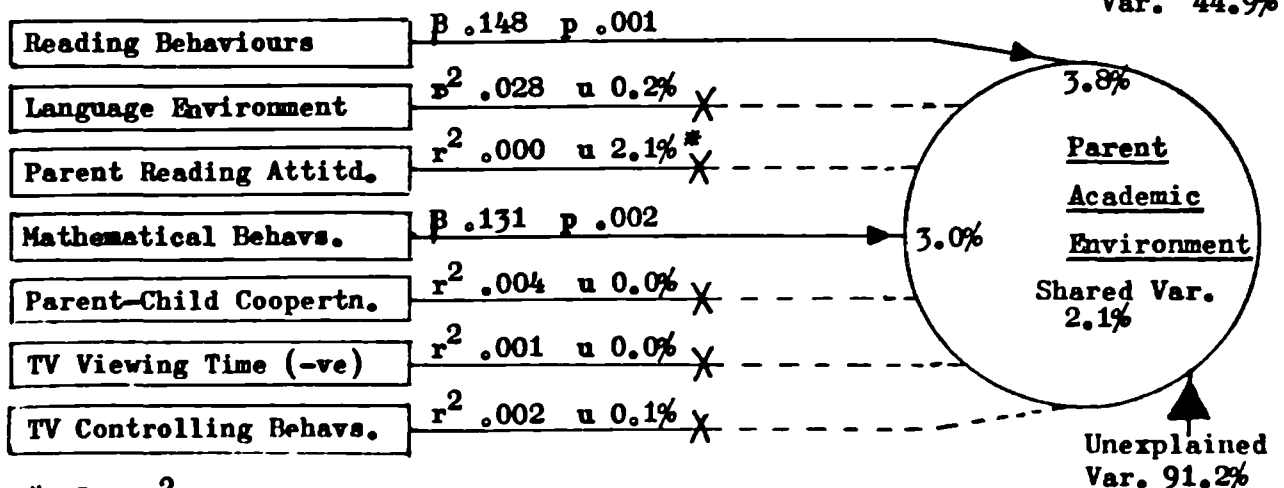
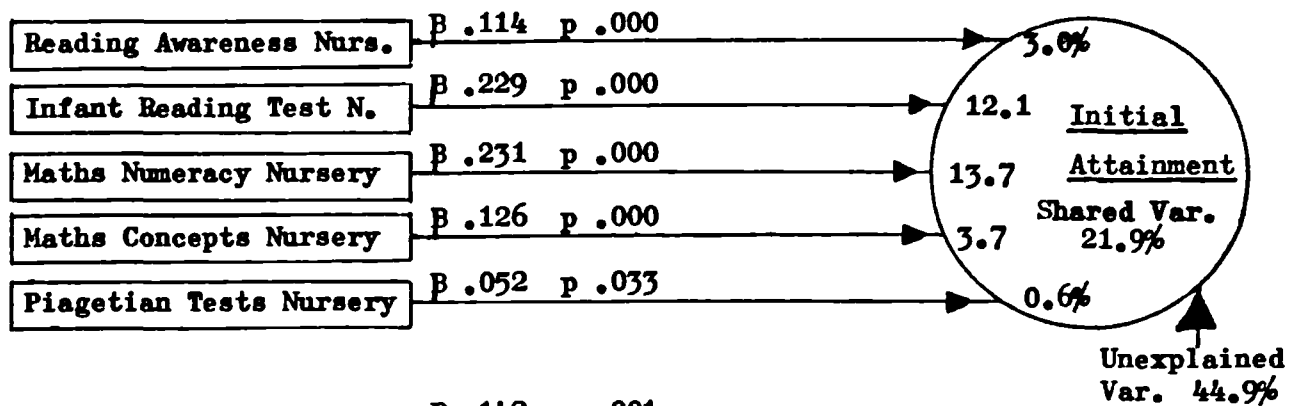
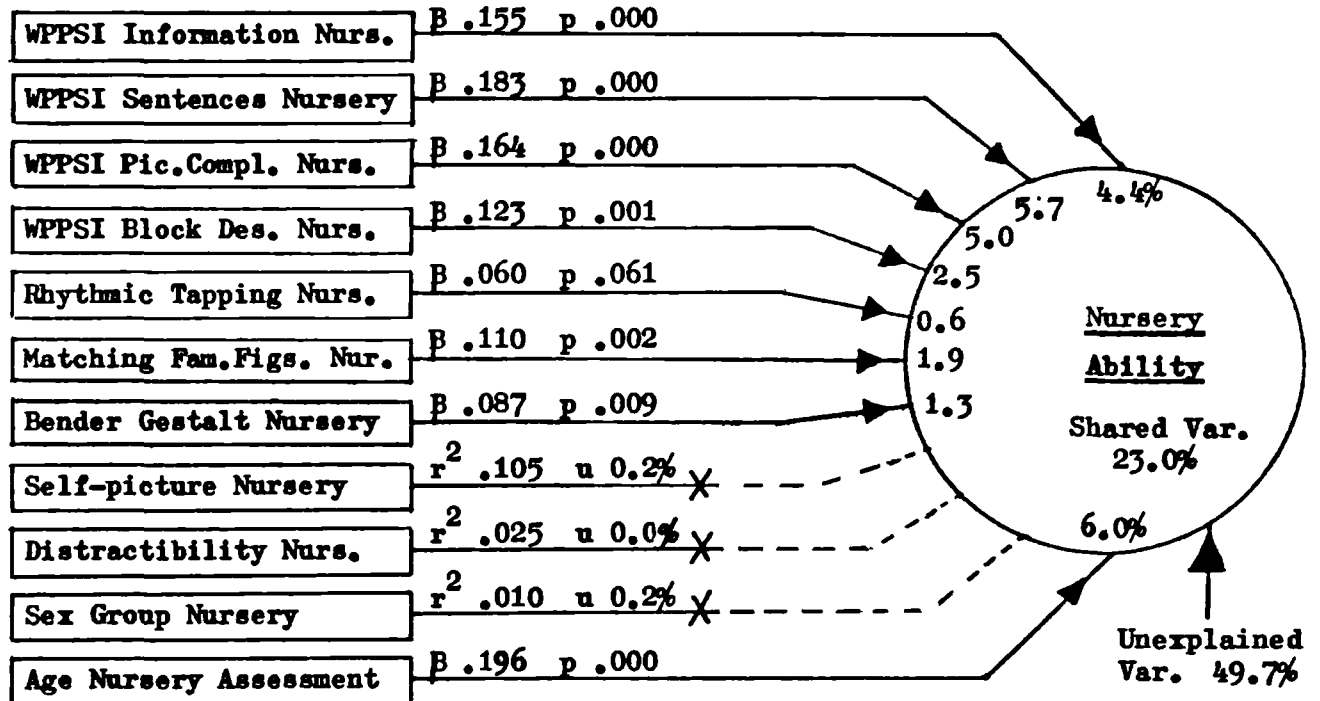
Parents who were working and therefore unable to attend meetings (28)

**Predicting:** Total academic attainment at the end of the study period (between 18 and 22 months after the initial assessments). The attainment consists of a weighted composite of three reading tests and three measures of early mathematical and Piagetian attainment. In all cases standardised scores were used. The weightings are defined in sub-section 6.831.

Creation of Latent Variables for Model 1

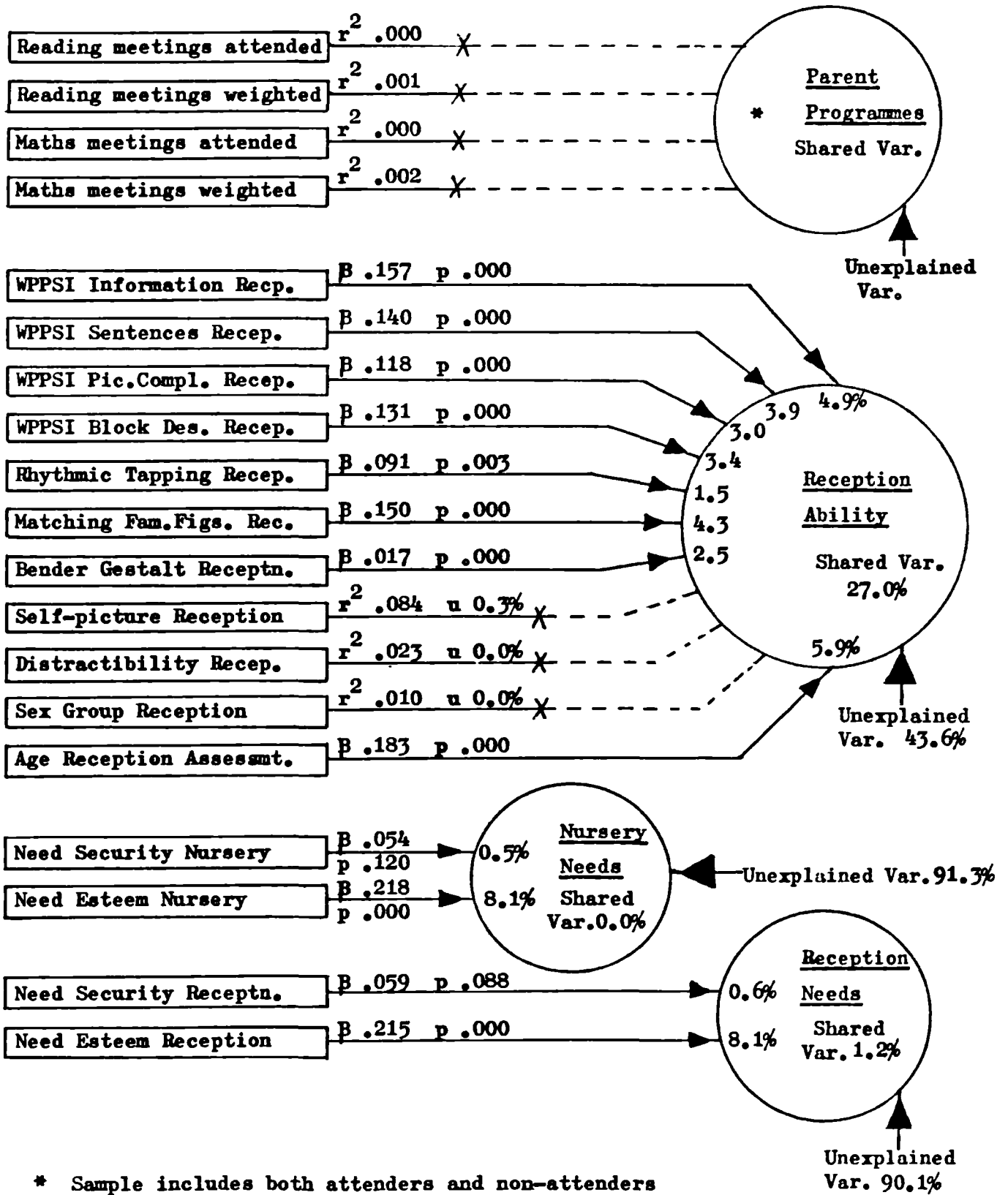
Sample: Total Disadvantaged (Base Model for Satellites) N = 129

Predicting: Total Attainment (Post-Test)



\* Low  $r^2$  and high u suggest multicollinearity effects

Creation of Latent Variables for Model 1 (continued)



### Model 1: Creation of latent variables

The regression of the outcome variable (Total Attainment) on the Nursery Ability set of variables shows that the four WPPSI intelligence scales used here, together with most of the meta-cognitive variables, contribute meaningfully to this equation and are therefore included in the creation of the latent variable itself. However the regression excludes Sex, Distractibility and Self-Picture. The same pattern appears when regressing outcome on to the Reception Ability set of predictors.

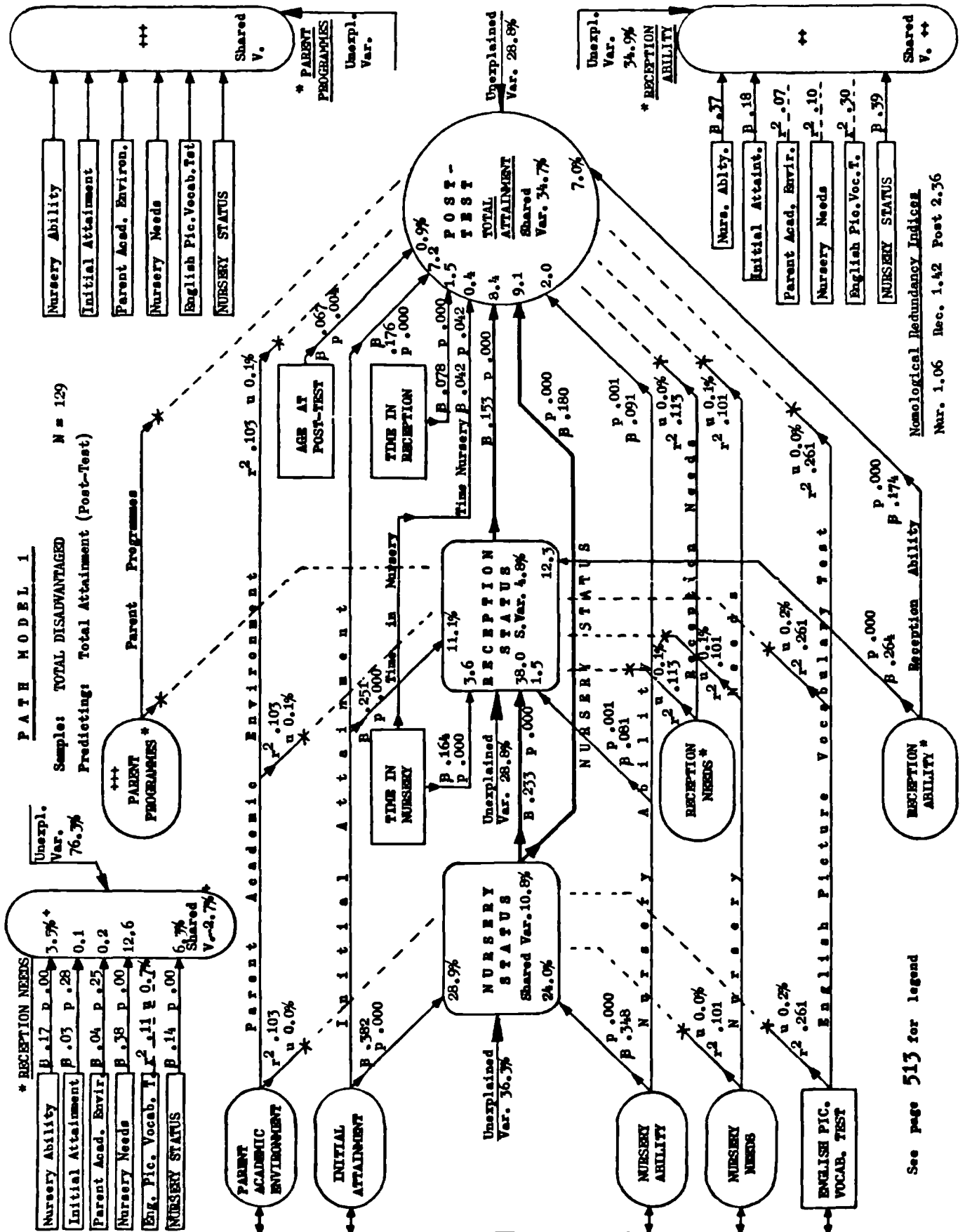
In general, Self-Picture (Draw-a-Person) has proved to be an unsatisfactory contributor in the analyses, despite its correlation with the outcome variables; it appears that it merely repeats the contributions made more effectively by the other meta-cognitive and cognitive variables. It is surprising that Distractibility is excluded, in view of the assumption that more disadvantaged children are likely to show more distractible behaviour. While their mean Distractibility scores are certainly lower (scored negatively) than those of the advantaged children, it appears that the measure of Distractibility does not in itself contribute to the total variance explained in this model, the squared correlation with outcome being only 0.025. As can be expected, the Reception Ability set on its own predicts slightly higher than does Nursery Ability on its own (56 to 50 per cent respectively).

In examining the creation of Initial Attainment it is interesting to note that despite the fact that Post-Test Total Attainment consists of three reading tests (weighted to total 50 per cent of that composite) and three mathematics measures (Numeracy, Concepts and Piagetian), it is Numeracy which makes the main Pre-Test contribution in this model, alongside the only reading test (I.R.T.) used at that stage. Maths Concepts makes very little contribution, and Piagetian even less.

Among the variables constituting Parent Academic Environment, only Reading Behaviours and Maths Behaviours make any meaningful contribution to outcome; the total prediction in the absence of any competing variables is only 9 per cent. It is noteworthy that one of the few attitudinal variables in the study, Parent Reading Attitude, which was discussed with parents longer than any other single item before being scored, has no predictive power in this or any other model, being completely overshadowed by behavioural variables. What adds to the interest of this point is that Parent Reading Attitude does correlate 0.51 with Parent Reading Behaviours, yet fails to have any correlation with academic outcome in this model.

The position with regard to the Parent Programmes variable is set out in the discussion of the construction of the path model.

Finally, the regressions on the two relatively simple Needs groups show that it is almost entirely Need for Esteem, rather than Need for Security, which contributes to outcome variance. It should be remembered that the Needs variables, as used in all the models, are negatively scored.



See page 513 for legend

Model 1: Construction of path model

The following points can be noted:

1. More than 70 per cent of outcome variance is predicted in this model. Given that the Pre-Test attainment measures were obtained approximately 20 months earlier, this is a reasonable prediction. It is important to recognise that at the Post-Test level half the predicted variance is shared, so that individual (unique) contributions at this level seem rather low. If in fact the shared variance was divided out in proportion to the unique contributions, individual predictors would seem more interesting than they do here.
2. The nomological redundancy indices for the first two levels in this model (Nursery or Pre-Test, and Reception) indicate very satisfactory regression equations for the model. The high percentage of shared variance at the final stage pushes up the third redundancy figures quite a lot, however.
3. The Attainment and Ability latent variables make approximately equal contributions to outcome at each level in the model.
4. The most noteworthy feature of the model is the fact that Parent Academic Environment makes a zero prediction when entered alongside other predictors. Its squared correlation with the outcome of 0.103 (correlation of 0.31) is only modest and clearly offers nothing that is not already being contributed by Initial Attainment and Nursery Ability. The implications of this finding, if supported by other studies, are very considerable. While it is known that parent behaviours in a disadvantaged population are less academic and less school-oriented than in an advantaged population, the fact that even the variance which is present makes so little contribution to outcome suggests that disadvantaged parents rarely rely on reading-related or maths-related activities in their daily interaction with their children.

The associated fact that Initial Attainment does nevertheless make a considerable contribution to outcome suggests that the disadvantaged children do have some early reading and mathematical concepts, but these may have come from the period they have already spent in the Nursery or - as was commonly reported to E during the interviews - from older siblings, rather than from the parent behaviours themselves.
5. Despite the relatively high squared correlation of 0.261 (correlation 0.51) between English Picture Vocabulary Test and the outcome, E.P.V.T. makes no unique contribution to outcome. Clearly whatever relationship there is adds nothing to the prediction made by Ability and Attainment. This is a puzzling finding. As reported earlier, the discovery that E.P.V.T. was eliminated from regressions on the subset of Initial Attainment variables (reading, maths and Piagetian)



led to the decision to enter it as a separate variable in the hope of improving the chance of obtaining an identifiable prediction from this vocabulary measure.

It has long been an article of faith (supported by much research) that the aspects of language represented by a well known test such as E.P.V.T. are an essential precursor of early academic attainment, and yet the performance on this test of word knowledge does not appear to bear out its prior importance. It may well be argued that the model itself, in particular the V-ridge regression algorithm, is at fault. In fact the Ordinary Least Squares regression for the same equation yields a negative coefficient for E.P.V.T. (despite the positive correlation) with a probability of .005 and a total (negative) contribution of 2 per cent to variance explained.

An alternative explanation to this problem of lack of prediction is that E.P.V.T. merely reflects a very limited range of verbal comprehension and is possibly biased against the disadvantaged. Yet even such a bias should highlight those upward mobile parents who are more academically oriented in their rearing behaviours, since all the evidence in this study has shown consistently that children from more stimulating or more advantaged homes have considerably higher scores on E.P.V.T. Thus, on its own, E.P.V.T. does serve as a predictor of later academic attainment.

It may well be argued that this aspect of language development is so closely related to early academic learning that in a model with academic attainment as the outcome variable, pre-test academic attainment not only mirrors early development of word knowledge but supplants it in statistical terms, since one could reasonably expect a closer relationship between pre and post academic attainment than between pre-test word knowledge and post-test academic attainment. Even that argument is brought into question by findings from the next model (focused on Advantaged children). The issue will be discussed further when Model 2 is examined.

6. Parent Programmes make no contribution to outcome (even in the original attempt to create a latent variable representing these measures). It may be asked why there should be no identifiable contribution, since both Disadvantaged Reading Groups and Disadvantaged Maths Groups (subsets of the Total Disadvantaged sample) show Parent Programmes to have a modest contribution to outcome? The reason is that within the total sample of 129 there are sizable groups who attended no programme meetings or who were working parents (and therefore unable to attend); even among the programme attenders those who attended reading meetings did not attend maths meetings, and vice versa. Thus this variable is not likely to be a meaningful predictor in a sample made up of four identifiably different groups.

7. Neither Nursery Needs nor Reception Needs make any independent contributions to outcome variance. Further discussion of these variables will be held over until the examination of later models.

8. Time in Nursery makes an important independent contribution of 3.6 per cent at the Reception level, and both Time in Nursery and Time in Reception make independent contributions at the Post-Test level. The large proportion of shared variance at the final level reduces the size of the contributions, but the probability figures indicate that there are no grounds for calling into question the predictive power of these variables. This is a particularly important finding, again subject as always to confirmation in other studies.

At a time when the value of nursery education is being called into question and when there have been sharp reductions in spending on this sector of education

- relative to other sectors - it is of some moment that analysis at the level of this study shows that Time in Nursery does make a useful independent contribution to academic attainment in the early years of school. The fact that measured Parent Academic Environment, E.P.V.T. and the two Needs variables all fail to make independent contributions to outcome variance in this model should highlight the importance of the finding that Time in Nursery does indeed contribute independently. It is interesting that its squared correlation of 0.184 is well below the figure of 0.261 for E.P.V.T.'s squared correlation with the outcome.

The creation of both the Time variables may be queried and indeed they presented a major conceptual problem, since it was necessary to remove from these variables the 'school effect' and to take into account the fact that the 'length of attendance' variables are closely linked to the child's Cohort Age (its absolute age in relation to a fixed datum line for all the children). School effect was eliminated by standardising each child's length of stay in Nursery, length in Reception and Cohort Age, on the means and standard deviations of its own school. The removal of Cohort Age from the 'length of attendance' variables is described in some detail in section 6.63. The rationale for the development of the two Time variables is set out in the same section. The grounds for accepting the credibility of the Time in Nursery variable are particularly strong, although Time in Reception also offers reasonable credibility.

9. The fact that Age appears at each level of the model (within the Ability groups at Nursery and Reception levels; and as a separate Age at Post-Test at the final level) offers the best evidence in favour of the contention that the two Time variables do indeed represent something much more fundamental than being surrogate age variables.

### Model 1: Subsidiary regressions

On the perimeters of the main path model are shown the unmodified initial regression of each Reception level variable on all the prior Nursery level variables. As explained earlier, the equations for these predictions have not been reduced to their final form, since they are used only as a general indication of how much variance in each Reception level latent variable might be explained by prior Nursery level variables.

About 23 per cent of the variance of Reception Needs is predicted by prior variables, mainly Nursery Needs, Nursery Ability (rather than Initial Attainment), and the Nursery Status variable itself. The question of whether Ability as such predicts to Needs levels (in other words, whether poor ability helps to contribute to later levels of Need for Esteem), or whether this relationship simply reflects a mutual interaction between the two characteristics, would need further analysis than could be undertaken in the time available for this study.

Since Parent Programmes do not feature as a predictor variable in this model, its prediction from Nursery variables was not assessed.

The prediction of Reception Ability yields an equation which would normally have merited refinement to eliminate the problems caused by poor predictors. (It was one of the limited number of occasions on which V-ridge regression failed to yield identifiable unique variances for each variable in a situation of serious multicollinearity.) However the size of the regression coefficients indicates, as expected, that most of the predicted variance of about 65 per cent comes from Nursery Ability and Nursery Status (in equal amounts), with Initial Attainment as a lesser contributor.

### Model 1: Overall findings

The following are the key findings which can be drawn from this model, subject as always to the limitations of the sample, the data themselves and the statistical algorithms on which the model is based.

- a. Parent Academic Environment for the Total Disadvantaged sample makes no independent contribution to Post-Test academic attainment. This is a disturbing finding, since it implies that the pre-reading and pre-mathematical 'behaviours' of disadvantaged parents in the home during the pre-school years do not contribute to academic outcomes, as measured late in the first year of school.
- b. Parent Programmes, as devised and provided in this study, make no contribution when viewing the Disadvantaged sample as a whole. This could however be expected given the fact that Total Disadvantaged is made up of four groups, two of which did not attend any Programme meetings.

c. Time in Nursery and Time in Reception both make useful contributions to academic attainment in the Reception year. This is potentially a most important finding, indicating that the time spent by disadvantaged children in the Nursery and Reception classes does add meaningfully and significantly to the variance in post-test attainment. It is particularly important in regard to the time spent in the Nursery, where the curriculum is focused more on general rather than on specific academic preparation for the work in the first year in school.

d. Ability and Attainment are equally strong predictors of post-test attainment for this group of children.

e. Language skills as defined by the English Picture Vocabulary Test do not make any identifiable unique contribution to post-test attainment, despite the recognised importance of word knowledge in early development and despite the fact that E.P.V.T. has a modestly strong correlation with final attainment in this sample. The reasons for this are complex and not readily apparent. The issue will be discussed further in subsequent models.

Model 2

Advantaged Sample ( 27 )

predicting

Total Attainment (Post-Test)

**Sample:** All the nursery class children at the advantaged school within the same inner urban education authority. As explained more fully in section 6.22, a few of these children's parents could not be interviewed. Data on two children were excluded after their parents said they would not be interested in attending programme meetings.

The 27 parents in this sample include:

Parents who attended reading group meetings (12)

Parents who attended mathematics group meetings (7)

Parents who agreed to attend group meetings but failed to do so (5)

Parents who were working and therefore unable to attend meetings (3)

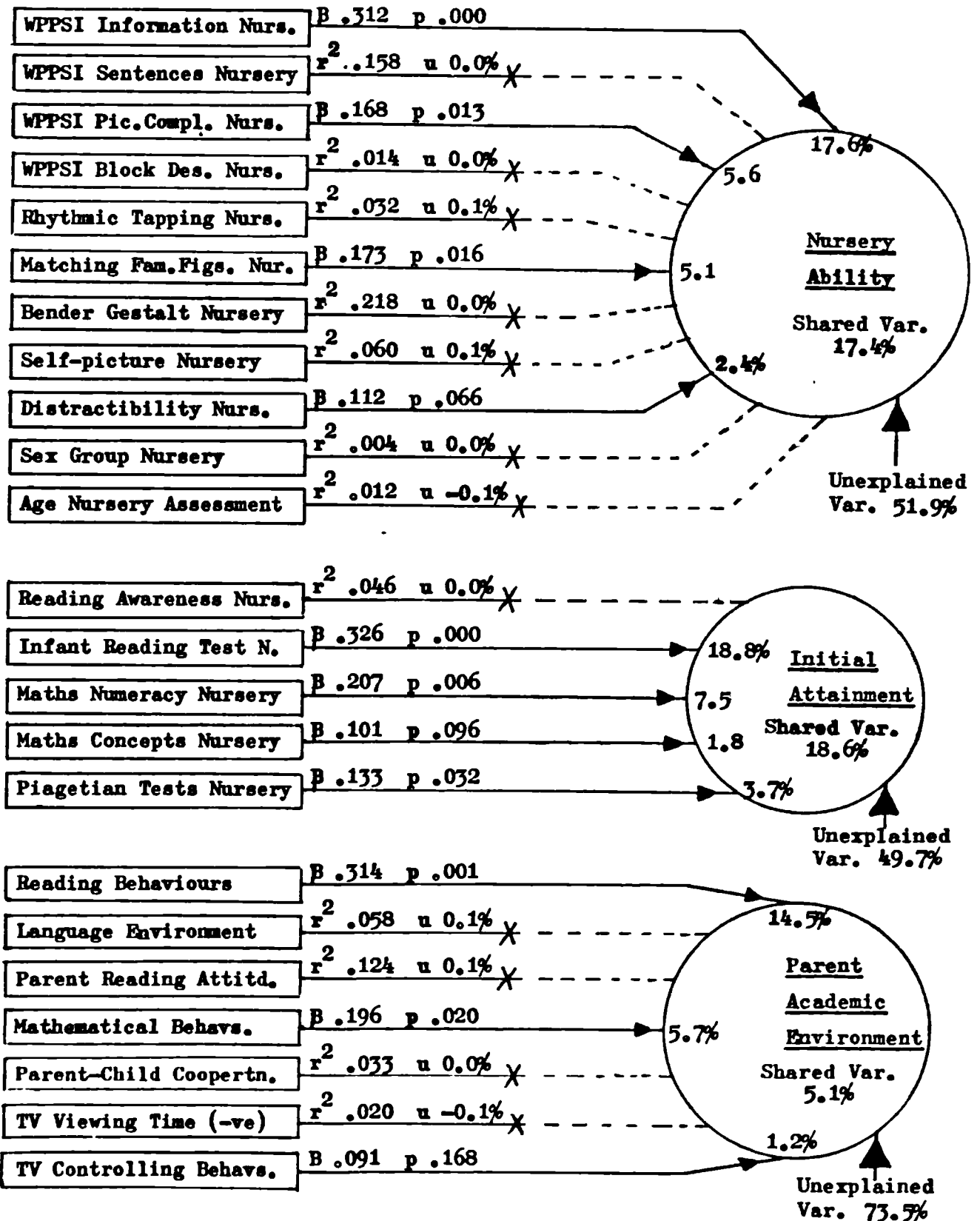
**Predicting:** Total academic attainment at the end of the study period,  
as defined in sub-section 6.831.

Creation of Latent Variables for Model 2

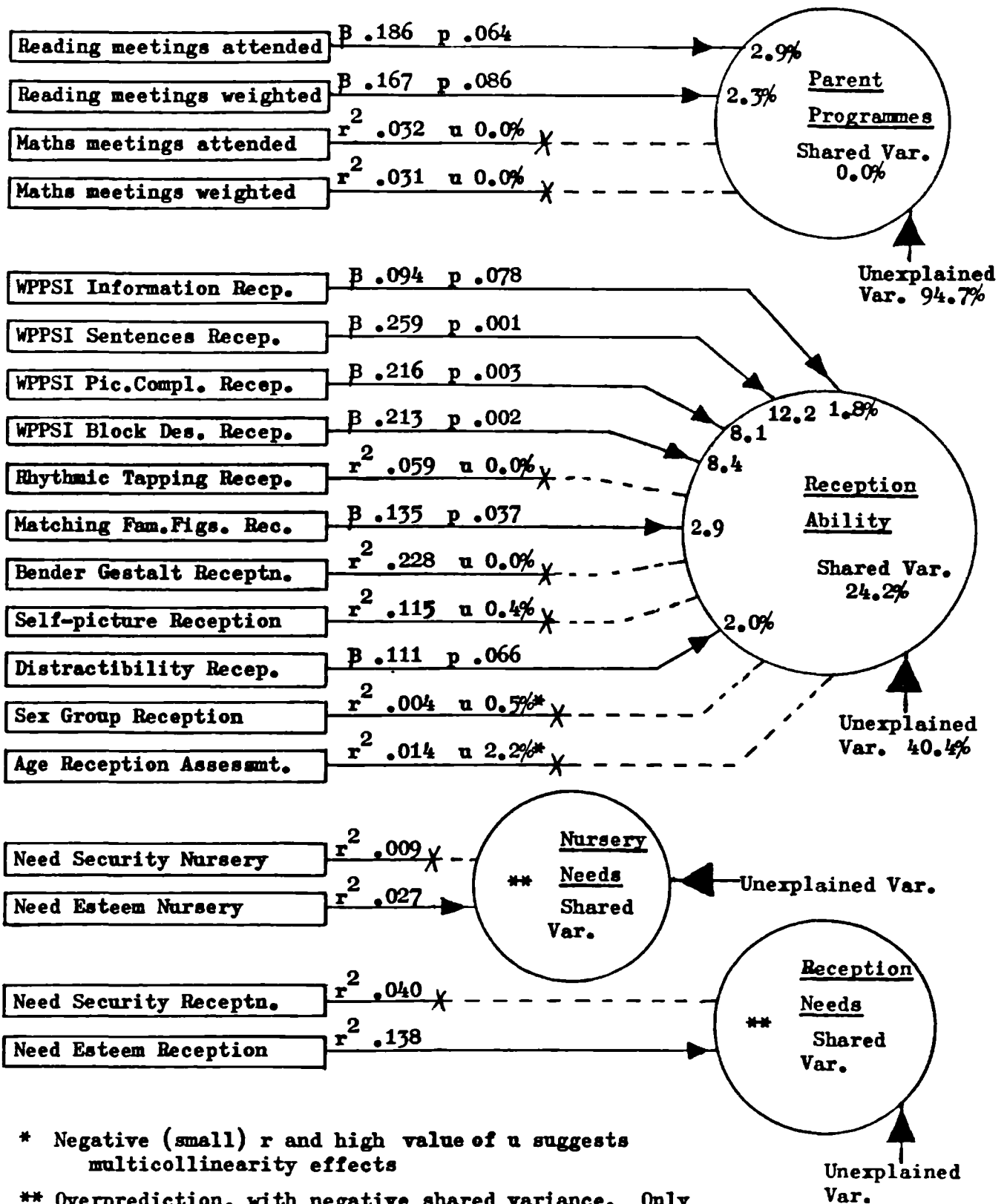
Sample: Advantaged

N = 27

Predicting: Total Attainment (Post-Test)



Creation of Latent Variables for Model 2 (continued)



\* Negative (small)  $r$  and high value of  $u$  suggests multicollinearity effects

\*\* Overprediction, with negative shared variance. Only the main contributor (Need for Esteem) used to form the latent variable

### Model 2: Creation of latent variables

Only relatively few of the Nursery Ability set of cognitive variables remain in the joint prediction to create this variable. WPPSI Information is the highest of these predictors. Surprisingly Distractibility appears as a modest co-predictor (although its probability is only .066). While the squared correlation of Distractibility with Total Attainment is 0.145 for the Advantaged Sample, it is only 0.025 for the Disadvantaged Sample, suggesting that the variation in Distractibility with advantaged children is greater or more meaningful than it is for their disadvantaged peers. Distractibility in Reception retains its predictive power (at the same probability level, coincidentally) within the Reception Ability battery.

The Reception Ability variable is more representative of the individual cognitive and meta-cognitive measures, although five predictors are still excluded by the regression results at this level.

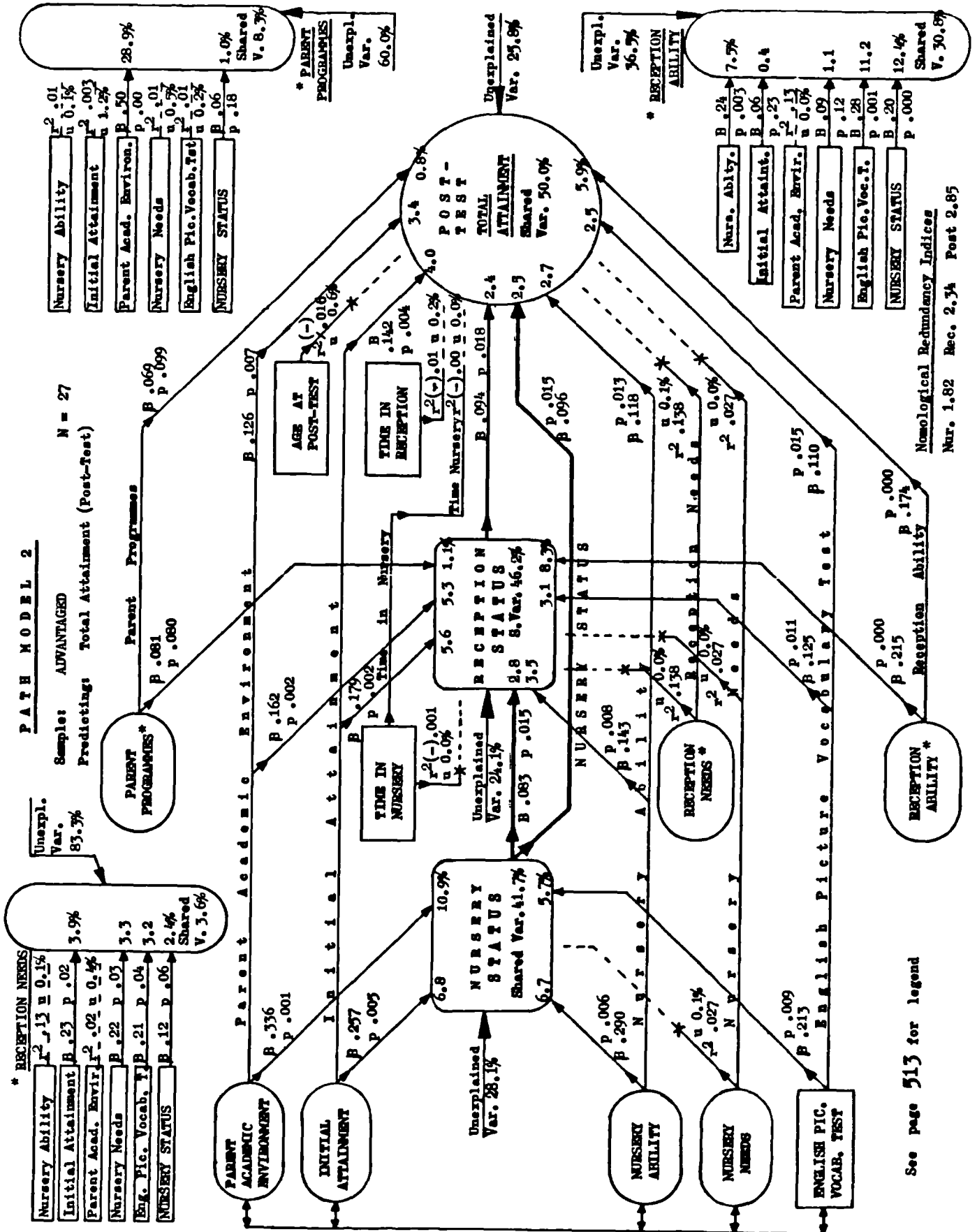
The Initial Attainment set of raw predictors shows a marked difference from the Disadvantaged set, with the Infant Reading Test being the principal contributor here and Piagetian tests contributing moderately strongly. Another surprise is that Reading Awareness makes no contribution (and has a low squared correlation with the outcome). This is almost certainly due to the likelihood that most of the advantaged children would be well aware of reading concepts (able, for example, to explain what words are and why should one want to read), so that the variation in this score would not be sufficient to serve any predictive function.

It is noticeable that the Parent Academic Environment set of variables contributes a total of 25 per cent to outcome variance (when on their own), compared to the equivalent total of only 9 per cent for the Disadvantaged set of Parent Academic variables on their own. While Reading Behaviours is the main contributing variable here, it is interesting that TV Controlling Behaviours also appears in this group, yielding a unique variance of 1.2 per cent. Although the coefficient probability of .168 makes it questionable as an important contributor, its presence accords with the finding of E in the interviews that advantaged parents were more likely to be aware of the subtleties of maintaining a sensitive control over the times and choices of their children's viewing.

The Parent Programmes variables on their own contribute over 5 per cent to outcome variance, although the regression equation allows only the reading meeting attendance variables to appear in the model.

The Needs groups of variables again show that it is mainly Need for Esteem, especially in Reception, which contributes to outcome variance.





Model 2: Construction of path model

The following points can be noted:

1. Approximately 74 per cent of outcome variance is predicted here. Even at the first (Nursery) level it is possible to predict nearly all of this total. It can also be seen that the total variance explained at Reception level is slightly higher than that explained at Post-Test level. In fact it is the correction formula, which takes number of cases and number of variables into account, that is almost entirely responsible for this anomaly. Since there are no new exogenous contributors added at Post-Test level, the uncorrected total predicted variance at Reception and Post-Test levels are virtually the same (0.824 and 0.822 respectively). The fact that Reception Status is added to the final regression means that the correction formula slightly reduces the total variance predicted at this level. (since Reception Status adds no new variance, but simply incorporates variance from variables at Reception and earlier levels).
2. The nomological redundancy figures are rather high at each of the three levels - especially the third. This suggests that the equations are not as 'valid' nomologically as one would have wished. Part of the reason is that there is a high shared variance within each major equation. While this phenomenon of a high shared variance may suggest that there is much overlapping of contributors, it also offers the interesting alternative hypothesis that the various skills, attainments and other characteristics of advantaged children are more integrated than is the case with disadvantaged children - and not simply a set of rather disparate characteristics.
3. The Attainment and Ability latent variables again make approximately equal unique contributions to outcome, although these limited totals (for example, 6.8 and 6.6 per cent respectively at Nursery level) together with the high shared variance already mentioned, again point to the possibility of a fairly high level of integration of the characteristics of advantaged children.
4. The massive unique contribution from the Parent Academic Environment (10.9 per cent at the Nursery level, with 5.3 and 3.4 per cent at subsequent levels), in a situation of high shared variances, indicates the very considerable importance of that environment for the attainments of advantaged children at this school. The contribution at the Nursery level is nearly twice that of any other variable. It is evident that the predictive power of the parent environment latent variable presents the most marked difference between the advantaged and disadvantaged models.

The question invariably arises as to whether this is a valid finding, in other words whether the measure of parent academic environment really reflects

what the interview protocol appears to reflect. If the protocol was basically faulty the measure would not have shown up the contribution of that environment to the advantaged children's performance - unless of course it were to be argued that the interview protocol operated in a grossly discriminatory fashion, failing to identify meaningful variance among disadvantaged parents while confirming it among advantaged parents.

The data themselves do not however indicate a serious absence of 'variance' for the disadvantaged sample. Comparative findings are as follows:

	Disadvantaged		Advantaged	
	Mean	s.d.	Mean	s.d.
Parent Reading Behaviours	7.77	3.47	8.56	3.62
Parent Maths Behaviours	3.43	2.02	4.70	2.52

While the variances of these behaviours in the disadvantaged sample are lower than those in the advantaged sample (12.04 compared to 13.10 for reading behaviours, and 4.08 compared to 6.35 for maths behaviours), the differences are quite understandable in relation to the lower means for the disadvantaged parents.

5. In contrast to its non-appearance in the disadvantaged model, English Picture Vocabulary Test scores predict a meaningful part of the outcome variance at all three levels in the advantaged model. This is in keeping with expectation but still does not explain why E.P.V.T. does not appear as a unique predictor for the disadvantaged sample. Even for the identifiably 'upward mobile' sample of working parents among the disadvantaged - whose children tended to perform better than the children of non-working disadvantaged parents - E.P.V.T. makes no identifiable contribution to outcome despite its squared correlation of 0.275 with that outcome. (The squared correlation with outcome for the Total Disadvantaged sample as a whole is 0.261.)

One possible clue to this problem is given by the fact that in the subsidiary regression (for the advantaged) of Reception Ability on to the Nursery variable set, E.P.V.T. makes a much higher unique contribution (11.2%) to Reception Ability variance than does even Nursery Ability (7.5%) - both of these predictions occurring within a situation of high shared variance. In contrast E.P.V.T. makes no contribution to the prediction of Reception Ability for the disadvantaged.

What this suggests is that E.P.V.T. is, paradoxically, much more closely linked to and integrated with the other characteristics of the advantaged children than it is for the disadvantaged. The paradoxical element lies in the fact that despite this close linking (and the high shared variance in most of the equations of the advantaged model), it still predicts independently when placed alongside the other predictors.

Why does this not occur in the disadvantaged sample? The evidence from these models is not conclusive but there are enough pointers to suggest that for the disadvantaged sample as a whole (though not necessarily for every dyad in that sample), word comprehension at this Nursery stage is not too closely related to the more fundamental conceptual skills measured by the WPPSI scales and other meta-cognitive tests, and is rather remote from the skills required in the early reading and mathematical measures used to assess total attainment. The accompanying table offers support for this hypothesis.

Table 39. Squared correlations between E.P.V.T. and Ability and Attainment

	Nursery Ability	Reception Ability	Initial Attainment	Total Attainment
Disadvantaged E.P.V.T.	0.469	0.300	0.399	0.261
Advantaged E.P.V.T.	0.538	0.659	0.757	0.574

As a general principle of language development it is clear that if a child has little or no word knowledge, she or he will be unable to cope with any attainment or other academic demands made by the school. But what may well be occurring is that at the generally lower E.P.V.T. levels of the disadvantaged children, the general word knowledge which they do possess is not yet closely linked to the specific word knowledge required for adequate performance in the cognitive and academic tasks given to them. This is of course not an absolute limitation, since the disadvantaged children often performed quite well on the various study measures.

What may be happening is that the word knowledge needed to cope with these specific conceptual and academic demands is to a limited extent already in their possession - it may have been learned in the Nursery class prior to the initial testing - whereas the disadvantaged children's total environment may not yet have given them the wider language skills which would normally be associated with these school-related tests.

Thus, for example, an advantaged child's richer language repertoire may normally include words such as 'word', 'set', 'centre', 'above/below'; there is likely to be a reasonable match between overall language usage and knowledge of conceptual and academically oriented words. In contrast a disadvantaged child will be more likely to grow up in a particularistic language environment (in the meaning given to this term by Bernstein, 1971), so that his or her knowledge and use of many words may be either lacking or fractionally different from the 'standard' meaning of these words as they are encountered in the school context. While the latter's repertoire may therefore be rather limited, the child may already have learned in the Nursery (in a relatively short time) the conceptual and academic words on which much early school work is of necessity based.

It is these considerations which may offer a theoretical explanation for the disparity in the predictive power of the E.P.V.T. across the very different social samples.

If the hypothesis can be confirmed it is a tribute to the influence of the Nursery environment that, despite the limited nature of disadvantaged children's general word skills these children have already been introduced to many of the linguistic concepts which they need to cope with school demands. It cannot be expected that any Nursery curriculum could add the many hundreds if not thousands of common words needed if such children are to reach the overall language level of their advantaged peers, in the space of a few school terms. The fact that most of them were able to cope with the essential concepts and with 'school' language when tackling the academic test measures is strongly suggestive evidence of how beneficial the Nursery environment can be for disadvantaged children.

6. Although the Parent Programme latent variable includes only the reading meetings variables, this measure makes a modest contribution at both the Reception and Post-Test levels (1.1 and 0.8 per cent respectively). While the probability levels are only .080 and .099, the fact that only 12 out of the 27 parents in this sample had any scores of reading meetings attended indicates quite a strong predictive power for the variable. Had the model been based only on the 12 reading programme participants the probability figures would have been a good deal lower. The finding on this contribution is in accord with E's experience of how enthusiastically the advantaged school parents attended the programme meetings and did the suggested work at home with their children - some of whom were reading well by the end of the programme period.

7. As with the Disadvantaged sample, neither of the Needs variables make any noticeable contribution within the Advantaged model. The low level of squared correlations with the outcome variable confirms that these variables are not important in this model.

8. The finding that Time in Nursery and Time in Reception make no contribution to the Advantaged model is in sharp contrast to the finding in the previous model. It suggests that advantaged children reach Nursery with so many skills of language, cognitive ability and early attainment - already developed in the home - that there is little the Nursery can do to add to those particular skills. Even the first (Reception) year at school may add only a limited amount to what most advantaged children already possess and may be so closely linked to the level of home stimulation that the latter remains the dominant predictor.

What this model does not examine, and this needs to be stressed very firmly, is the extent to which advantaged children benefit socially and culturally from the group environment and the considerable variety of creative and cultural activities offered by the Nursery. There is little reason to doubt that the social

mixing and the various activities available within a Nursery are of considerable benefit to advantaged children. What this study did examine was the specific contribution of the parent programmes and other Nursery related variables to the development of early academic skills.

What should also be emphasised is that this data was based on the performance of children within a single Nursery class - compared with the disadvantaged data which amalgamated the material from children in five Nursery classes. On the other hand there was nothing in the nature of the particular advantaged Nursery class to suggest that it did not offer the fullest possible range of Nursery activities and opportunities for mixing and interacting between children and between children and adults. If anything, there were fewer social problems to contend with in that Nursery than in the Disadvantaged Nursery classes, and thus more opportunity for creative and other activities. There were also fewer disruptive episodes - resulting from the activities of a minority of emotionally troubled children - such as tend to occur not infrequently in schools sited in the seriously deprived parts of the Metropolitan authority where this study took place.

9. Age appears to make no independent contribution in this model, neither within the Nursery and Ability latent variables nor as a separate Age at Post-Test. The squared correlations of the age variables with the outcome are close to zero - 0.012, 0.014 and 0.016, with  $r$  negative. Questions in regard to the age variables have been discussed in section 6.61, where it was pointed out that the relatively limited relationship of age with academic outcome has been diluted still further in the advantaged sample by the fact that there were several unusually brilliant children at the youngest age level in this Nursery sample.

What may also be the case with an advantaged sample such as this is that at an early age it is the level of child ability, the parent environment and other environment-related variables such as E.P.V.T. which make the major contributions to early educational attainment within a particular age group, almost regardless of the specific age distribution across that group. Thus the age finding may not be altogether atypical of advantaged children in general, although the expected attenuated relationship has clearly been reduced to meaningless by the presence of the handful of unusually bright young children in this class.

#### Subsidiary regressions

Of the three equations shown on the perimeter of the main model, that of Reception Needs, regressed on to the set of Nursery latent variables, shows little of real interest.

The Parent Programmes regression is, however, most interesting since it shows that the only real predictor of attendance at Parent Programme meetings (in

this case reading meetings) is the Parent Academic Environment itself. In other words the variance in the Parent Academic Environment, as assessed in the original interview prior to the start of the Parent Programmes, predicts nearly three-quarters of the 40 per cent explained variance of the Parent Programmes variable. Part of the remaining shared variance is possibly also linked to the parent environment variable. When it is remembered that less than half the sample have any scores on the reading meetings which make up the Parent Programme variable, the closeness of the association between original home environment and attendance at meetings is emphasised.

It should thus be recognised that, for this sample at least, attendance at programme meetings is largely predictable from the academic 'strength' of the home environment; it may be argued that E's programmes were in reality accelerating what was in any case occurring in these homes, namely the teaching of reading to the children. This is an important finding to which reference will be made later. At the same time it should be emphasised that the finding of an additional contribution by Parent Programmes to the outcome variance, beyond the considerable amount which is already contributed by Parent Academic Environment, is a genuine and independent contribution and not simply a surrogate for the parent environment.

Mention has already been made of the fact that English Picture Vocabulary Test is the strongest exogenous predictor of the third mid-test group, namely Reception Ability; at 11.2 per cent the unique prediction by E.P.V.T. is well above the 7.5 per cent uniquely predicted by Nursery Ability.

### Overall findings

The following are the key findings which can be drawn from the model, subject to the usual caveats in regard to the sample, the data and the analytical algorithms used here. For this model an additional caveat is that the findings are based on data obtained from children and parents at a single school.

- a. The patterns of relationships within the disadvantaged and advantaged models differ very considerably. Either it has to be concluded that the advantaged school is atypical and not like other advantaged schools, or that the differences between the population groups are genuine and possibly of considerable importance.
- b. The Parent Academic Environment is the most important single contributor to the advantaged model, while it does not appear at all as a predictor in the disadvantaged model. It is not unexpected that parent environment should be so prominent in the advantaged model, but there are worrying educational policy implications of the finding that despite the power of the statistical algorithm developed for this study, there is no evidence of any unique contribution by parent environment to the Reception year academic attainment of disadvantaged children.
- c. Parent Programmes do predict modestly to outcome variance, though only at a

level of  $p .080$  and above. Given that the sample includes just 12 who attended reading meetings out of a total sample of 27 (only reading meetings contributed to this regression outcome), it does indicate an acceptable contribution. On the other hand the extremely close predictive association identified between Parent Academic Environment and Parent Programmes reduces the value of the programme for this sample. While this fact suggests that the Parent Programmes were essentially intensifying what parents were already doing for their children, it should be stressed that the regressions in the main model do identify clearly distinct contributions to outcome by Parent Academic Environment and Parent Programmes separately.

d. The chosen measure of academic attainment shows no identifiable contribution of Time in Nursery and Time in Reception for these children. As emphasised earlier, this measure takes no account of social, creative and other forms of development which are also integral to the Nursery experience.

e. Ability, Attainment and E.P.V.T. all predict to outcome, at approximately the same levels.

f. In general the evidence from this model points to a far more integrated development of advantaged children than is in evidence from the disadvantaged model. The high shared variance, coupled with the identifiable contributions from expected contributors such as attainment, ability and verbal comprehension, and the variance added by the parent environment, all present a portrait of a sample whose varying skills are being developed rapidly but in harness with each other. This is perhaps a case where powerful statistics can reflect and confirm what is often intuitively thought about the differences in the early development of advantaged and disadvantaged children.



Model 3

Disadvantaged Working Group ( 28 )

S A T E L L I T E M O D E L

predicting

Total Attainment (Post-Test)

**Sample:** The nursery class children of those disadvantaged mothers who were working outside the home in the day-time and were therefore unable to attend the parent programme meetings. These mothers were a sub-sample of the Total Disadvantaged Sample described in Model 1. All 28 mothers in this group said that if they had not been working they would have liked to have taken part in the programmes.

**Predicting:** Total academic attainment at the end of the study period, as defined in the introduction to Model 1.

**Satellite Model:** The principles followed in developing this satellite model have been set out fully in sub-section 6.825. Briefly, the creation of the latent variables and of the status variables in the satellite path model has been based on the set of regression parameters derived for the Base Model (the Total Disadvantaged Sample). This means that the underlying structure of the satellite model is closely bound to that of the Base Model; it facilitates comparison but also means that the parameters of the satellite model are not totally independent of the Base Model.

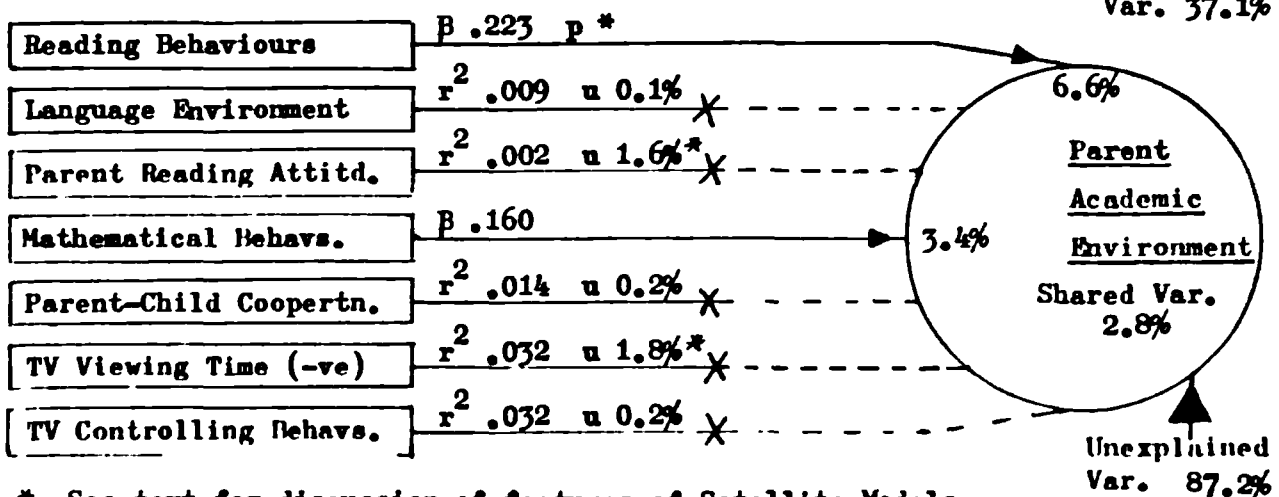
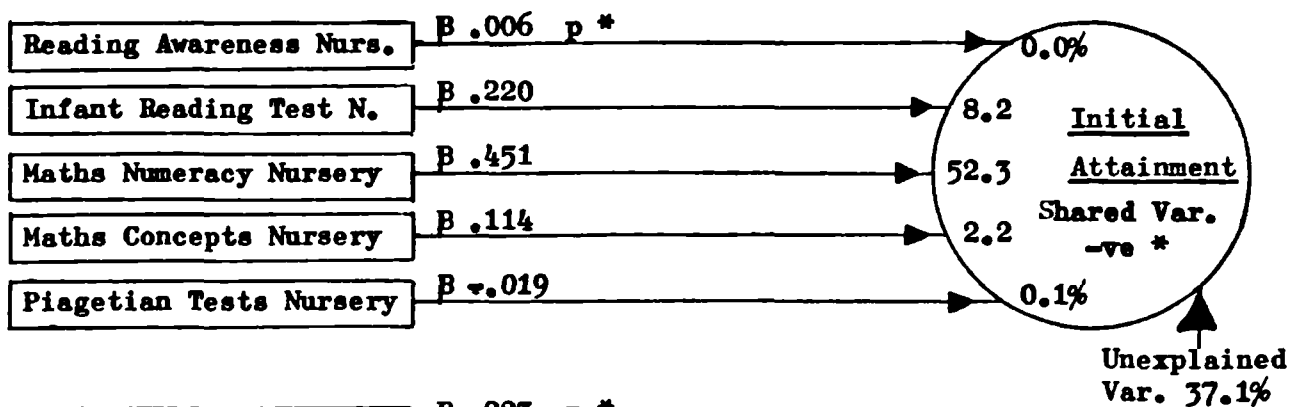
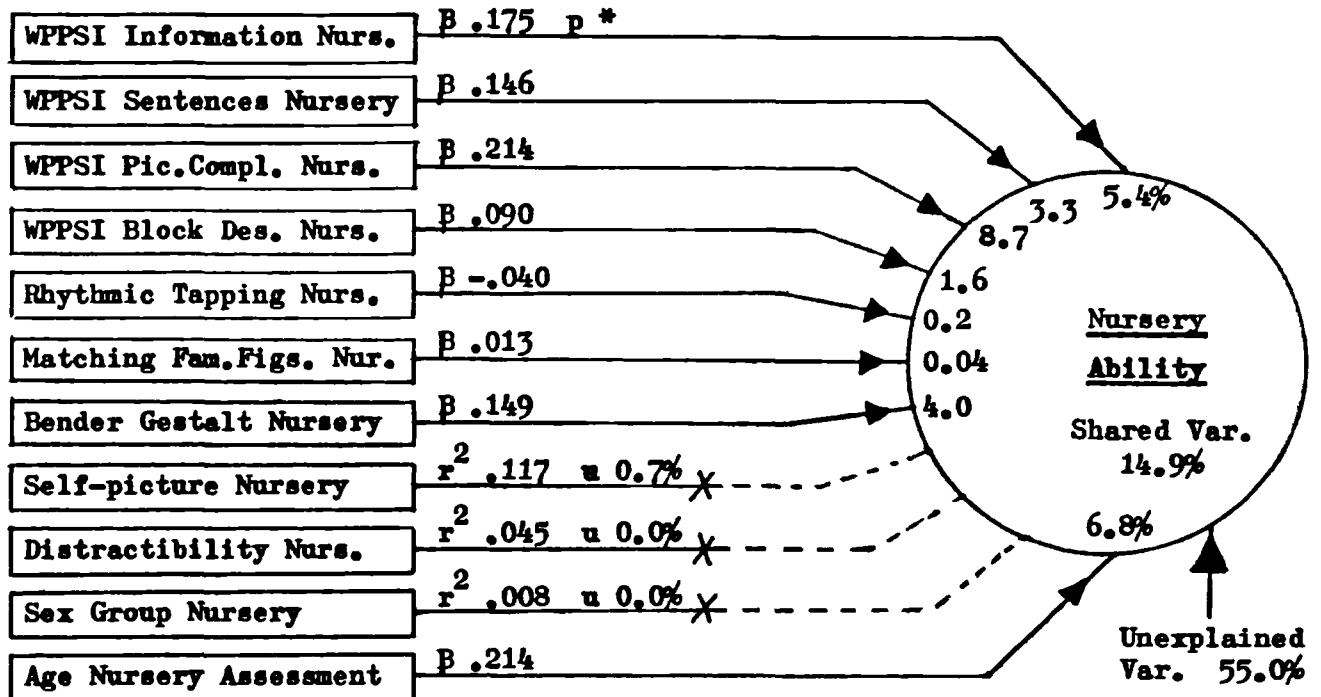
Where there are clear departures between the structure of the Base Model and that indicated by the parameters of the satellite (Disadvantaged Working Group) model, these are pointed out in the description which follows. It should again be emphasised that all the regression parameters described in the following pages are unique to the Disadvantaged Working Group. Only the formation of the latent and status variables is common to the Base Model and all the satellite models.

Creation of Latent Variables for Model 3

Sample: Disadvantaged Working Group (Satellite Model\*)

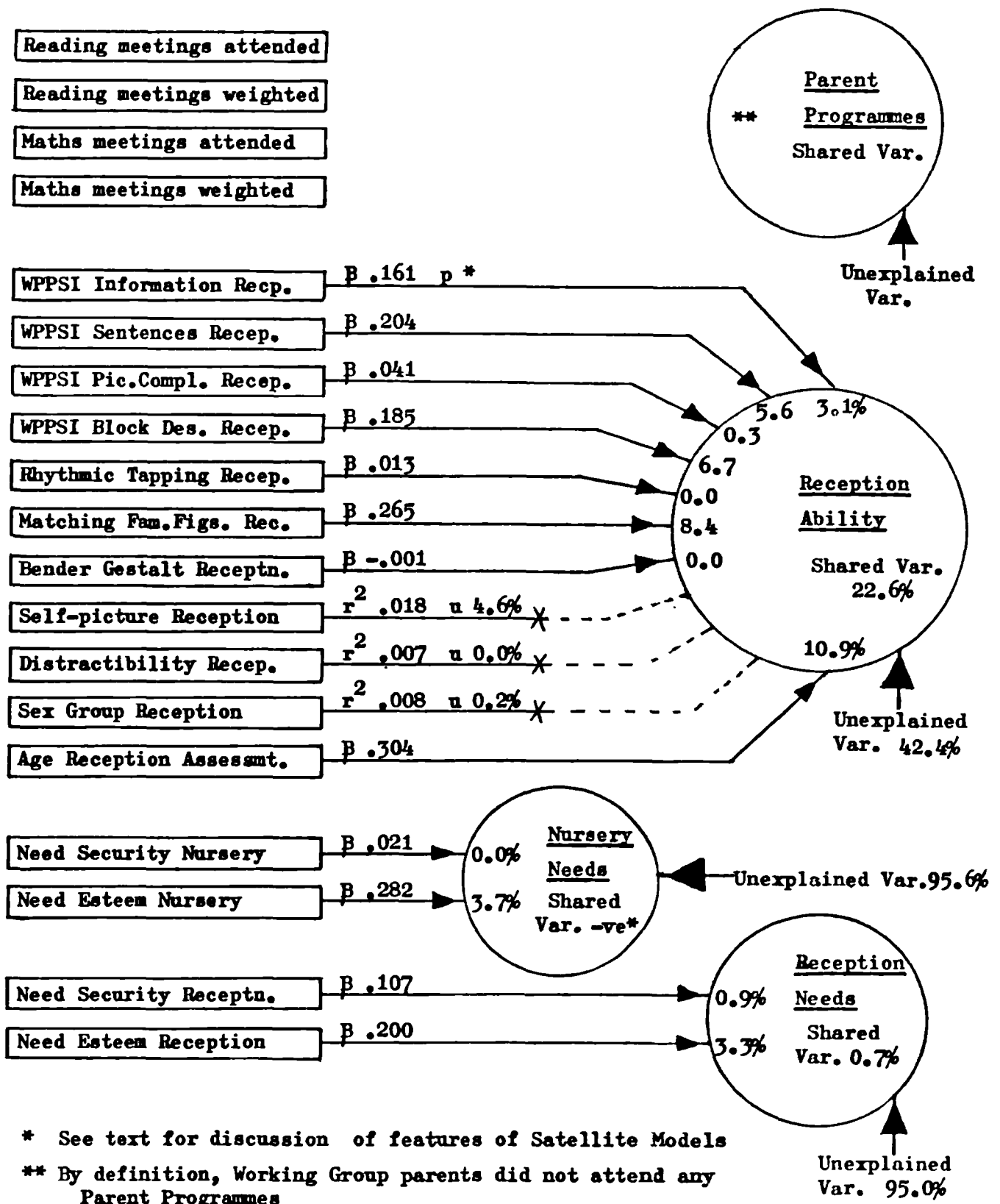
N = 28

Predicting: Total Attainment (Post-Test)



\* See text for discussion of features of Satellite Models

Creation of Latent Variables for Model 3 (continued)



### Model 3: Creation of latent variables

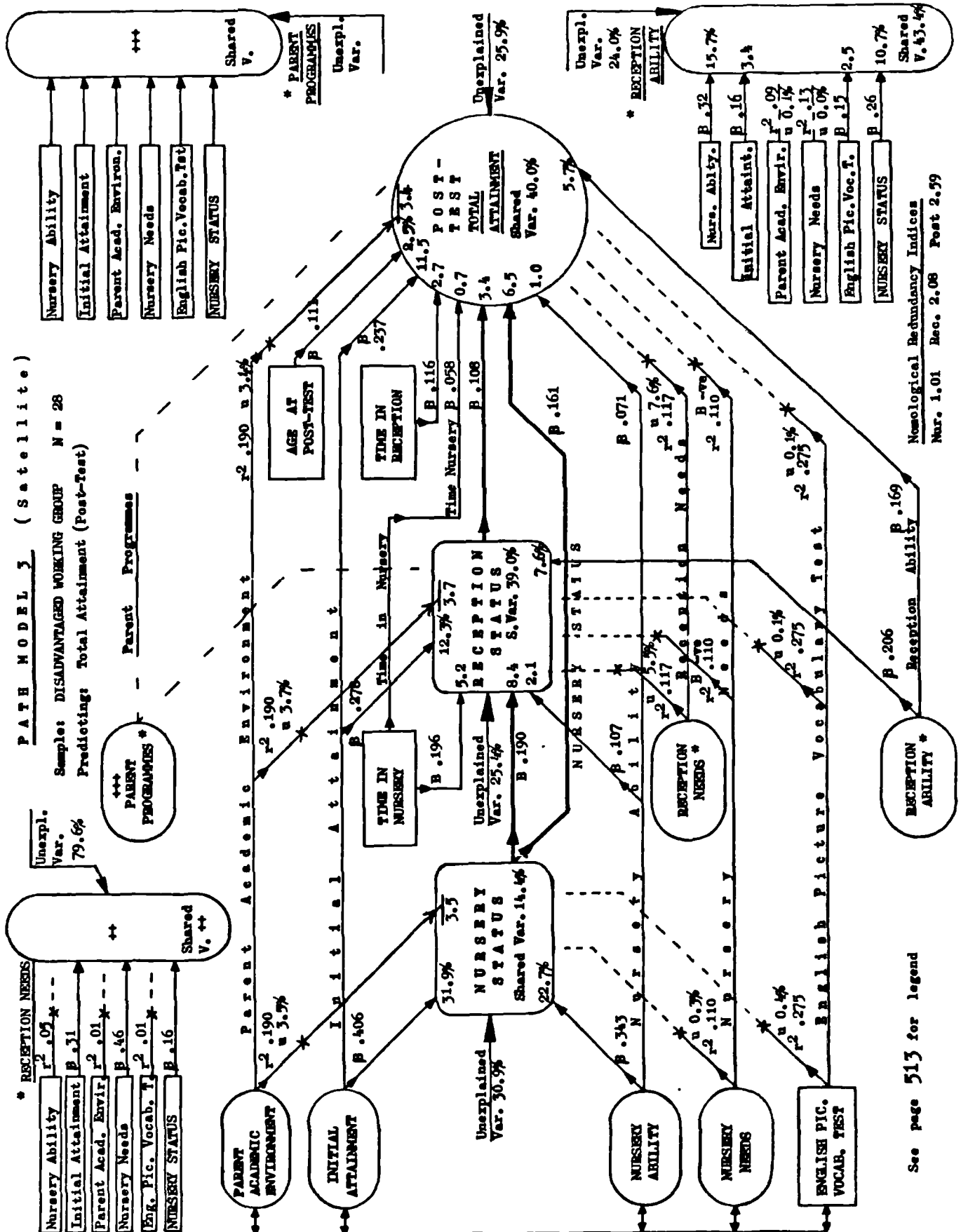
With the exception of Bender Gestalt and Self-Picture the parameters for the two Ability latent variables in the Disadvantaged Working Group and the Total Disadvantaged Sample (Base Model) do not vary greatly. Bender Gestalt contributes moderately at Nursery level but not at all at Reception level, for the Working Group. Self-Picture makes an apparently noticeable contribution of 4.6 per cent at Reception level, but this is based on an insignificant squared correlation of .018 and can almost certainly be disregarded on the grounds that the phenomenon is related to multicollinearity within what is quite a small sample.

There is a much larger and more noticeable difference in the parameters of the set of Initial Attainment variables. Here Maths Numeracy at the Nursery level predicts over 50 per cent of the outcome variance for the Working Group, but under 14 per cent in the Base Model. As the Working Group's Maths Numeracy mean is only a little higher than that of the Total Disadvantaged Sample, the explanation for this difference appears to be rather obscure.

There is little to remark about the parameters for the other latent variables, excepting that multicollinearity in the full equation for Parent Academic Environment incorrectly highlights some predictors with low squared correlations with the outcome. This statistical weakness is probably due to the small size of the sample in relation to the very small predictive power of the seven parent variables.

In total, it can be said that while there are sizable differences in some of the regression parameters (between those of the Base and Working Group models), none of the Base Model's latent variables have excluded any powerful contributors which are found only in the Working Group. A model derived from the Working Group would certainly have weighted initial Maths Numeracy more highly (at about twice its Base Model weighting), but other than this added weighting of the Initial Attainment variable the latent variables created for the Working Group would not have differed greatly from those of the Base Model.

Differences which are fairly noticeable are in the size of the variances explained by the latent variable groupings in each of the models. The Working Group's Initial Attainment set on its own predicts 63 per cent of Total Attainment (Post-Test), compared with only 55 per cent for the equivalent prediction in the Total Disadvantaged Sample; Parent Academic Environment on its own predicts 13 per cent for the former model, compared with 9 per cent for the latter (Base Model); on the other hand the two Needs variable sets on their own each predict only around 5 per cent of outcome variance for the Working Group, compared with around 10 per cent each for the Base Model.



See page 513 for legend

### Model 3: Construction of path model

The following points can be noted, in comparing the regression parameters for the Disadvantaged Working Group (Satellite Model) with those for the Total Disadvantaged Sample (Base Model).

1. The total outcome variance explained by the Working Group model is slightly larger than that of the Base Model, while the shared variance in the former model is considerably larger than in the latter.
2. Nomological Redundancy indices are, as could be expected with the constraining conditions of the satellite model, somewhat higher than for the Base Model, but not seriously so.
3. The Attainment and Ability latent variables each contribute handsomely to the outcome in the Working Group, but an important difference is that for the Working Group the Initial Attainment contributes a good deal more, at every level, than does ability; for the Total Disadvantaged Sample the Attainment and Ability variables contribute approximately equally.

This is the first important difference identified between the Base model and the Working Group model. While the attainments of the Working Group are on the whole a little higher than those of the Total Disadvantaged Sample, what seems to have been pinpointed here is that the Working Group may be more 'attainment oriented' than the Base sample. Other evidence already cited appears to support this hypothesis, namely that the Working Group are more upward mobile and achievement-oriented.

4. Parent Academic Environment appears as a modest contributor in the regression for the Working Group. Since this variable is not absorbed by either of the status variables (whose construction is dependent only on the parameters of the Base Model), Parent Academic Environment continues to hold the same predictive power at each of the three stages of the Working Group model.

What is important about this finding is that, whereas for the Total Disadvantaged Sample the Parent Academic Environment makes no contribution whatever to outcome variance - when entered in competition with other predictors - the same variable does predict in the Working Group model. This adds weight to the possibility of identifying the Working Group parents as upward mobile, since the predictive power of Parent Academic Environment is a particularly strong feature of the Advantaged model.

It should be noted that on the diagrams for the satellite path models, variables which do not appear as predictors in the Base Model but which appear to make a meaningful contribution to the satellite models (such as Parent Academic Environment in this case) have their potential contributions noted with an over or under-

line, within the status variable or outcome variable perimeters. The line indicates that such variables would be present as predictors in an independent model, but are not in fact incorporated in the creation of the Base Model's status variable.

5. E.P.V.T. remains stubbornly outside the model, despite the slight increase in squared correlation (from 0.261 for the Total Disadvantaged Sample to 0.273 for the present sub-sample). It may be hypothesised that although the Working Group appears to be more 'achievement-oriented' than the Base sample as a whole, the level of verbal comprehension measured by E.P.V.T. continues to make little or no independent contribution to outcome variance.

6. Parent Programmes clearly cannot feature in this model, since the Working Group were by definition unable to attend any programme meetings.

7. The predictive behaviour of Nursery Needs and Reception Needs in the path model needs particular attention because of the apparent importance of these latent variables. Each variable set on its own predicts about 5 per cent of the outcome variance. For a comparison the same latent variable sets on their own predict 9 and 10 per cent of variance in the Total Disadvantaged Model. At the Nursery level in the Working Group model the Nursery Needs variable makes only a very low prediction of 0.3 per cent when competing with the other predictors, and is accordingly left out of consideration. The picture is thus little different from that experienced in the Total Disadvantaged model, where Nursery Needs has no predictive power whatever.

However at the Reception level, within an initial equation of nine predictors, the presence of Reception Needs forces the competing Nursery Needs predictor into a negative prediction (i.e. a negative regression coefficient, although adding positively to the predicted variance), and in consequence the Reception Needs variable itself assumes a high positive predictive strength. The effect becomes even more exaggerated at the Post-Test level, in an initial equation of 12 predictors, when Reception Needs appears to be making the second highest prediction at that level. The following points can be made about this anomalous behaviour.

(i) The path model regression equations from which the contributions of the two Needs variables were taken were the initial equations into which all the predictors had been entered simultaneously. On the basis of these equations poor predictors are normally removed and the equations are then re-run, if necessary followed by further removal and a final run of the ultimate equation.

(ii) Since this was a satellite model it was not possible to run separate sets of reduced equations for the Working Group, as the equations had to follow the Base Model reduction sequence. Thus it was not possible to assess what the effect might have been had Nursery Needs alone been removed from the Reception and Post-

Test level equations.

(iii) All the evidence from the Total Disadvantaged and Advantaged models indicates that neither of the Needs variables have any useful predictive power in those models; it is only for particular sub-groups discussed in later models that there is a sufficiently high prediction at the initial (latent variable) stage to have any continuing effect when entered alongside other predictors in the path models.

(iv) The behaviour of Nursery Needs at the Nursery level (unique prediction 0.3 per cent) and the evidently close predictive relationship between Nursery Needs and Reception Needs, as shown in the peripheral regression of Reception Needs on the Nursery level variables, offer no evidence that Reception Needs could be more than a marginal predictor within this path model if entered in the absence of Nursery Needs.

(v) An examination of the means of the Needs variables and their squared correlations with the outcome variable shows little indication that the Working Group children have a serious level of Needs, compared to other groups. Some of these parameters are as follows:

Sample	Standardised means		$r^2$ with outcome variable	
	Nur. Needs	Rec. Needs	Nur. Needs	Rec. Needs
Total Disadvantaged	0.0	0.0	0.101	0.113
Working Group (Disadv.)	0.196	0.318	0.110	0.117
Programme Groups (Disadv.)	0.035	-0.108	0.162	0.248

Clearly the Working Group has a better mean on this negatively scored variable (indicating a lower level of Needs) than do the combined Reading and Maths Programme groups. The squared correlations with outcome are also at the same low level as they are for the Total Disadvantaged Sample.

Thus, on both comparative and statistical grounds it seems reasonable to treat the apparently high prediction of Reception Needs in this satellite model as a statistical anomaly which could not be resolved within the satellite situation. None of the evidence cited here offers any grounds for treating Reception Needs as a meaningful predictor in this model.

8. Time in Nursery and Time in Reception both make useful contributions to variance in this model - more so, in fact, than in the Base Model. The question arises as to whether Working Group children benefit more by their attendance at Nursery (and Reception) than do the Total Disadvantaged Sample children as a whole. The fact that there is a higher contribution of Parent Academic Environment for the Working Group may militate against such a hypothesis. However an examination of the mean scores for Time in Nursery and Time in Reception shows that, relative to the Base Model children as a whole, Working Group children are on



average about one-fifth of a standard deviation above the mean for the Total Disadvantaged Sample for these two specially constructed variables (and about one-quarter s.d. above the mean for non-working group children within the Total Disadvantaged Sample). While higher mean values do not of themselves indicate any reason for the higher contribution of the Working Group 'time' variables to outcome variance, higher values do suggest a longer attendance at Nursery class relative to the child's cohort age. Only a more detailed study of individual cases could offer definitive evidence on this issue. The higher predictive value may well be due to a greater capitalisation on Nursery experience by the Working Group children.

9. The Age variables, both when incorporated in the Ability variables and as Age at Post-Test, contribute slightly more within the Working Group than they do within the Base Model sample.

#### Subsidiary regressions

There is not much of note in these regressions. Overall the Working Group model parameters appear to be little different from those of the Base Model. There is no strong reason to question the tentative findings reported in previous paragraphs. One interesting finding is that E.P.V.T. does appear as a predictor of Reception Ability, in the peripheral regression of that variable on the Nursery level variables in this sample.

#### Overall findings

The following are the key findings which can be drawn from the Disadvantaged Working Group model, subject to the usual caveats in regard to the sample, the data and the analytical algorithms used here. For this model an additional limitation is that it is a satellite model, so that although all its regression parameters are based only on the sample itself, the creation of the latent and status variables was constrained to imitate that of the Base Model and thereby facilitate direct comparisons.

a. In a number of ways the Disadvantaged Working Group sample appears to have characteristics somewhere in between those of the Total Disadvantaged Sample and the Advantaged Sample. There is also more shared variance in the Working Group model than in its Base Model; shared variance is a particularly strong characteristic of the Advantaged Sample model. On the other hand there are still a number of features of the Disadvantaged model which are common to both the Base Model sample and the subsidiary Working Group.

b. The greater importance of Attainment, as compared to Ability, in the Working Group model is a feature which distinguishes it from both Total Disadvantaged and

Advantaged samples.

c. Parent Academic Environment appears as a meaningful predictor in the Working Group sample; although well below the figure for the Advantaged Sample, it is at least present, whereas it is absent from the Total Disadvantaged Sample. (For reasons explained earlier, probability figures are not provided in support of satellite model parameters.)

d. E.P.V.T. remains a non-predictor in this model, although it does feature as a small but meaningful predictor of Reception Ability, in a peripheral regression. Again this bears some modest similarity to the Advantaged Sample.

e. Time in Nursery and Time in Reception show meaningful contributions in the Disadvantaged Working Group model, above those in the Total Disadvantaged model. This and related evidence about the mean values of these variables suggests that the Working Group makes marginally better and earlier use of Nursery class facilities than does the remainder of the Disadvantaged sample. In regard to these (Time) variables the Working Group differs sharply from the Advantaged Sample; in the latter neither of the Time variables makes any independent contribution to post-test academic attainment.

f. The evidence in this model pointing to a possible contribution of Reception Needs to outcome variance is dubious and can with reasonable safety be disregarded as a statistical artefact deriving from the limited flexibility of satellite models.

g. In general this model offers evidence that the Disadvantaged Working Group parents and their children do differ in important respects from the Total Disadvantaged Sample of which they form a part. The differences are of both educational and sociological importance.

Model 4

Disadvantaged Reading Groups ( 48 )

predicting

Reading Attainment (Post-Test)

**Sample:** All the Nursery class children of the 48 mothers who attended the different reading programmes at five disadvantaged schools within the same inner urban education authority. The programmes consisted of eight one-hour sessions held at fortnightly intervals over a period of four months. The parent attendance at programme meetings was recorded and used to create the Parent Programmes variable in the path model.

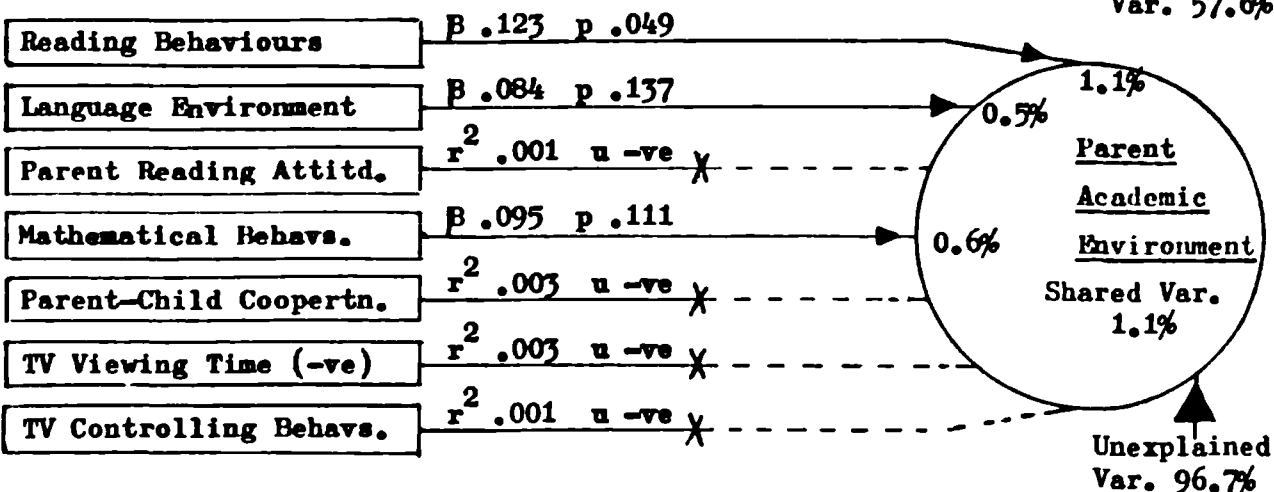
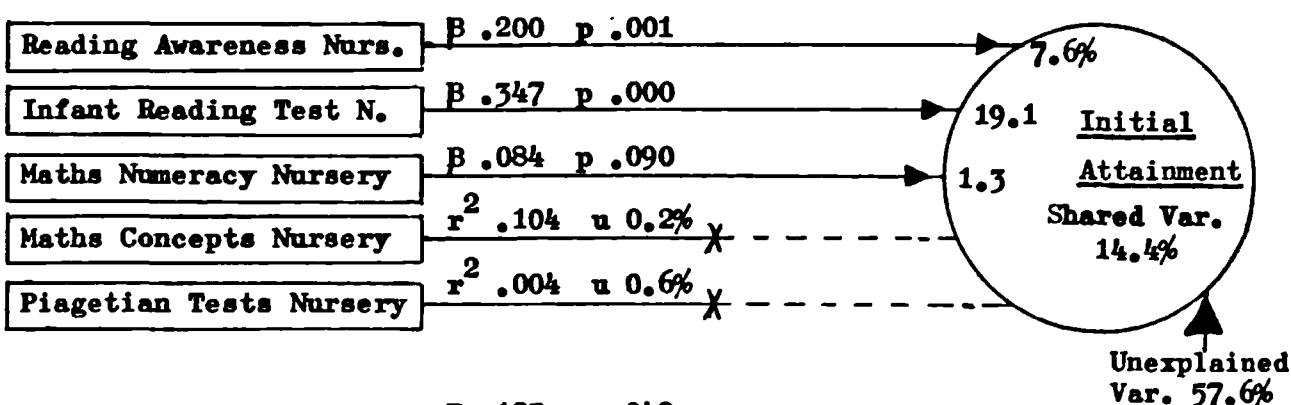
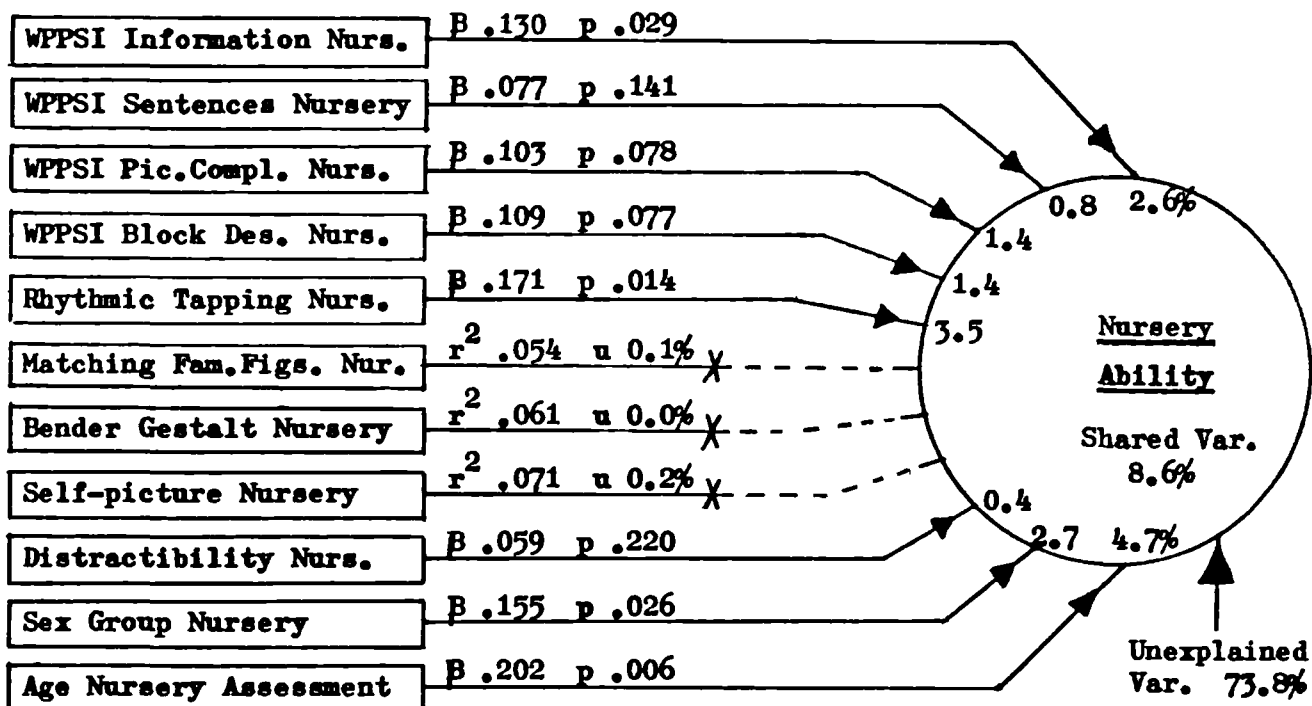
**Predicting:** Reading Attainment at the end of the study period, this variable being a weighted composite of the Infant Reading Test, the Southgate Reading Test and the Daniels and Diack Sentence Reading Test, using standardised scores. The weightings for each test are defined in sub-section 6.831.

Creation of Latent Variables for Model 4

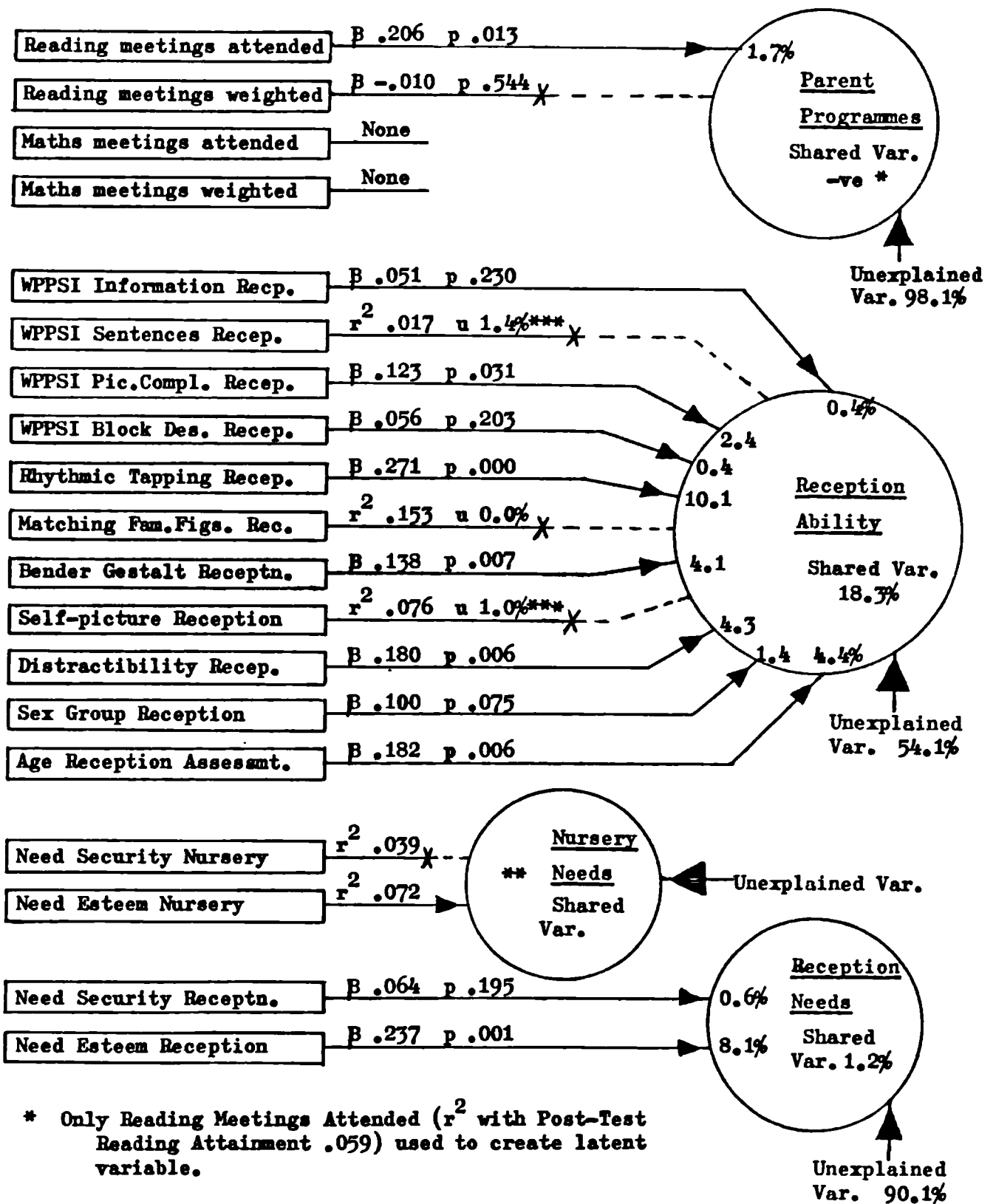
Sample: Disadvantaged Reading Programme Groups

N = 48

Predicting: Reading Attainment (Post-Test)



Creation of Latent Variables for Model 4 (continued)



\* Only Reading Meetings Attended ( $r^2$  with Post-Test Reading Attainment .059) used to create latent variable.

\*\* Overprediction, with negative shared variance. Only the main contributor (Need for Esteem) used to form the latent variable.

\*\*\* Regression coefficient negative, although correlation positive, suggested multicollinearity effects.

#### Model 4: Creation of latent variables

The regression equations used for the construction of the Nursery Ability and Reception Ability latent variables indicate, rather unexpectedly, that the WPPSI Sentences scale contributes very little to outcome variance (reading attainment) in either the Nursery or Reception set of WPPSI tests; at the Reception level WPPSI Sentences makes no independent contribution. This suggests that short-term memory for meaningful sentences may not be an important prior skill in the learning of early reading.

While Matching Familiar Figures and Self-Picture make no contribution (both measures having low squared correlations with the outcome), Bender Gestalt makes a useful 4 per cent addition to outcome variance at the Reception level, but none at the Nursery level, suggesting that the variation in this perceptual skill at Nursery age does not discriminate strongly among pre-readers. Another variable which makes almost no contribution at Nursery level, but again a useful 4 per cent at Reception level, is Distractibility. Given that Distractibility seldom appears in these models, its contribution within a model focused on the learning of reading suggests that the teaching of reading in Reception is indeed affected by the level of concentration even among disadvantaged children.

A particular feature of the two Ability latent variables is the contribution of Rhythmic Tapping; the Reception measure of this skill explains 10 per cent of outcome variance within the Reception group. The high prediction mirrors the finding of De Hirsch et al (1966) that auditory perceptual integration is a strong predictor of early reading.

As can be expected, Infant Reading Test and Reading Awareness form the bulk of the prediction of Initial Attainment; however Maths Numeracy does add a small contribution.

The total explained variance of Parent Academic Environment on its own is only 3.3 per cent, with Reading Behaviours, Mathematical Behaviours and Language Environment each contributing minimally. It is noteworthy that (Parent) Language Environment does not appear as a predictor in any other model; even here it makes only a 0.5 per cent contribution (with  $p$  of .14).

Reading Meetings Attended contributes well within the Parent Programmes variable, but as this meetings attended variable is the only meetings predictor taken into the latent variable, further comment on this will be reserved for the discussion of the path model.

The variables contributing to Nursery Needs make only a small contribution to outcome variance; only Need for Esteem is retained here. The contribution of the two Reception Needs variables is much larger, with the combination predicting nearly 10 per cent of variance.



Model 4: Construction of path model

The following points can be noted:

1. The percentage of outcome variance explained in this model is less than that explained in many of the other models, amounting to only 50 per cent at Nursery level; however the increase in variance explained during the 20 months of the study is interesting, as it reaches a total of 67 per cent at the Post-Test level. When comparing these figures with the much bigger and earlier predictions in the Advantaged Sample (72 per cent at Nursery level, rising to 74 per cent at Post-Test level), the Disadvantaged Reading Groups prediction pattern suggests that for the latter children a large part of the variation in reading skills resides in the children's school experiences - possibly including motivational variables arising out of the curriculum as well as many other unmeasured school variables. The low percentage of shared variance at each level suggests that the contributors to early academic skills in these disadvantaged children are not highly integrated as they are in the Advantaged Sample, for example.
2. The nomological redundancy indices for each of the three major stages of this model are particularly sound, none of the indices being above 1.50. This indicates that the equations at each level have a strong nomological validity and do not rely on a number of highly overlapping and relatively poor predictors.
3. Initial Attainment contributes a good deal more to the model than does Nursery Ability, and in fact the latter disappears as a predictor after the first level regression, being replaced by Reception Ability. However unlike the previous model, where Initial Attainment for the Disadvantaged Working Group continued throughout as a major predictor, considerably more powerful than either Nursery or Reception Ability or even the two combined, the early importance of Initial Attainment for the Disadvantaged Reading Groups subsides rapidly. Again this points to the likelihood that the parents and children of the upward mobile Disadvantaged Working Group were more attainment oriented from the outset than were these non-working group parents in the Reading Programmes.
4. Parent Academic Attainment makes no unique contribution to outcome variance in this model. Its low squared correlation of 0.094 with the outcome, coupled with its low prediction of little over 3 per cent in the regressions creating the latent variables, emphasise the negligible contribution of parent home behaviours to post-test reading attainment.

If this is the case it may be asked how it can occur that Initial Attainment (comprising in this case mainly Infant Reading Test scores and Reading Awareness scores) can contribute as much as 22 per cent to outcome variance at the time of initial testing in the Nursery (in the path model)?

The explanation for the predictive power of the Reading Awareness variable



may well be sought in the experience given in the Nursery classes from the outset, where disadvantaged children do become aware of the purpose of books and hear of concepts such as 'words' and 'reading', in many cases possibly for the first time in their lives.

The explanation of the predictive power of the Infant Reading Test scores is more complex. The test itself comprises a variety of early reading skills, such as writing ones own name, identification of simple environmental words and letters, and success with the Marie Clay test which requires knowledge of where to start reading in a combined picture and story book, and how reading progresses along and down the lines. Even in homes where there is little emphasis on early reading children are often taught to write their names, particularly by older siblings already at school. Likewise children may learn obvious environmental words such as Bus Stop and STOP when they accompany their parents on shopping expeditions. A little knowledge of letters, especially of the letters comprising their own names, is also often learned from siblings. The skills assessed in the Marie Clay test items may well have been picked up in the first weeks or months in the Nursery class.

5. Again the English Picture Vocabulary Test makes no independent contribution, even though in this model it is reading variables alone which form the outcome. Here E.P.V.T. has a squared correlation of only 0.10 with the outcome. This finding, coupled with the findings in earlier models about the performance of E.P.V.T., suggests a number of tentative conclusions.

It is recognised that the test is a measure of spoken word recognition rather than of verbal comprehension as a whole, and clearly it does not assess expressive language skills. At the same time the fact that the level of word knowledge - as assessed by this test - does not predict to early reading suggests that the skills needed for the preliminary stages of reading in a disadvantaged sample may be related more to motivational factors and learning abilities in the children, coupled with the teaching skills of the teachers, than to the amount of language experience gained by the child up to that time. This is not to suggest that language experience is not of fundamental importance to high level reading skills; but it does point to the possibility that early word knowledge is not the overriding factor in pre-reading experience. There is of course a minimum level of word comprehension which is needed before any school learning can take place. But given this foundation, what may well occur is that initial reading develops from a complex of skills and interests and as the basic reading skills are mastered so does the level of language comprehension become increasingly important for the fuller exploitation of those initial skills.

It may thus be asked whether the emphasis of the Nursery curriculum should not be expanded. Its present commitment is, in brief, to creative and social experience, to the development of fine and gross motor manipulative and percep-

tual skills, and to an immersion in language as a hypothesised precursor to the learning of reading in the first school year. The present study suggests that the learning of early reading skills could also be attempted at Nursery level for those children interested in learning, without waiting for the completion of the language widening experience which is fundamental to much educational thought about the function of the Nursery class.

Clearly the development of language is an ongoing necessity, in Nursery as in all succeeding school classes. But there is little hard evidence to suggest that the development of early reading should be delayed until some arbitrary level of language attainment has been reached; not unexpectedly, the arbitrary level is usually that attained by the child at the time of transfer to Reception class. In contrast, the evidence of this and other disadvantaged sample models in the present study suggests strongly that while there is a moderate correlation between word knowledge (as measured by E.P.V.T.) and early reading, this relationship has no predictive value for the early reading attainment of disadvantaged children.

A combination of early reading tuition with early language experience may therefore provide an even more interesting and useful Nursery environment for disadvantaged children than that which they already receive, with the talk surrounding the explanation of early reading skills being as useful for further language development as are any of the other aspects of language experience in the Nursery classes of today.

6. One of the most important findings of this study is that the Parent Programmes variable does predict modestly at both the Reception and Post-Test levels (1.7 and 1.6 per cent), with probabilities of .029 and .025 respectively. This appears to be a reasonable result for eight one-hour sessions with the parent groups over a four-month period. As the variable itself comprises the number of meetings attended, one may ask whether fortnightly parent meetings at the Nursery throughout the period of the child's stay in that class might not contribute even more strongly to early attainment, provided the parent programmes were well structured?

7. As in earlier models, neither Nursery Needs nor Reception Needs make any unique contribution to explained variance.

8. Time in Nursery and Time in Reception make a higher contribution to this model than to any of the other independent models in the study; only the Disadvantaged Working Group (in a satellite model) shows a higher contribution of the 'Time' variables to outcome variance. The unique predictions of 4.2 and 3.7 per cent variance respectively are of considerable importance, given the competing power of the Ability and Attainment variables within the same model. The probability levels of .002 are also quite satisfactory. These figures again empha-

sise that both the Nursery and Reception periods are of major importance in the development of early reading skills among disadvantaged children.

9. The various Age variables make a modest contribution to this model, with no prediction being of particular note. The age contribution within each of the Ability groupings is a little over 4 per cent.

#### Subsidiary regressions

The subsidiary regression of Reception Needs on the Nursery level variables yields a prediction of about 28 per cent of variance, most of this coming from Nursery Needs, Nursery Ability and the composite Nursery Status variable. The relatively high contribution to Reception Needs by Nursery Ability (6.6 per cent, compared to 6.1 per cent from Nursery Needs) is quite interesting.

The regression of the Parent Programmes latent variable on the Nursery level variable set is an important one. It has already been shown in the Advantaged Sample (Model 2) that the Parent Academic Environment is the only large contributor to Parent Programmes in that model, suggesting that the better the home's academic environment the better the attendance at Programme meetings. In the case of this disadvantaged sample, with 48 parents in the reading programmes at five schools, the parent academic environment makes no contribution whatsoever nor does it have any meaningful correlation with the attendance scores of these parents at the Programme meetings. None of the other variables make any contribution to Parent Programmes within the regression equation, the highest squared correlation being 0.08 between E.P.V.T. and Parent Programmes.

The importance of this finding is that the parent programmes can be seen as moderately effective within a highly disadvantaged population, regardless of whether the individual parents are already showing 'desirable' academically-oriented behaviours (such as regular reading to the children) in the rearing of these children.

The regression of Reception Ability on the Nursery level variables show that nearly all the predicted variance comes from Nursery Ability and the composite Nursery Status (the latter consisting of Initial Attainment and Nursery Ability).

#### Overall findings

The following are the key findings which can be drawn from the model, subject to the usual caveats in regard to the sample, the data and the analytical algorithms used here.

a. The parent programmes were effective for the disadvantaged reading groups, regardless of the home academic environment. The variance predicted by pro-

gramme attendance, although modest, was approximately one-third of the variance predicted by Time in Nursery and Time in Reception.

b. Early attainment is more important than early ability in predicting post-test reading for these disadvantaged children, but Reception Ability thereafter appears as the stronger contributor to final attainment.

c. Within the Ability latent variables Rhythmic Tapping ability, particularly that measured at the start of the Reception class, appears as a strong predictor of post-test reading attainment.

d. The Parent Academic Environment makes no independent contribution to the outcome of post-test reading attainment within the model as a whole; even on its own, with no competing predictors, this variable makes only a small contribution of just over 3 per cent. This adds strength to the evidence from the Total Disadvantaged Sample (Model 1) about the minimal predictive value of current 'academic' behaviours in disadvantaged homes. At the same time there may be other motivational and cultural values in the home which have not been assessed and which may contribute significantly to reading outcome in Reception class.

e. The Time in Nursery and Time in Reception variables are each important contributors to post-test reading, suggesting that the experience and curriculum in both these classes are of much value in establishing early reading skills. While the contribution of the Reception experience is to be expected, the model also offers statistical confirmation of the importance of Nursery schooling in the early education of children.

f. The evidence from this model suggests that word knowledge, as measured by E.P.V.T., does not contribute to the prediction of early reading attainment. This suggests that the belief that considerable language development is needed before the learning of reading skills is possible, is not necessarily correct. What may well be the case is that reading development may be promoted best by the simultaneous fostering of reading skills and language development within the Nursery curriculum, with the growth of the former skills being a powerful early motivator for the linguistic and all round academic development of the child.

g. The percentage of variance predicted in this model is somewhat smaller than in any of the previous models, suggesting that what is learned in the Nursery and Reception classes, possibly in combination with other unmeasured variables, contribute a good part to the variance in post-test reading. Had it been possible to assess and quantify the variation in teaching methods and quality and the variation in the time given to the learning of reading, the predictive power of the model may have been higher.

h. The fact that there is little shared variance at any level in the model suggests that the skills of these children are not well integrated; this is in

accord with finding (g) above, since there is clearly much room for the addition of other unmeasured contributors to outcome variance, such as the input of the teachers themselves. The fact that Parent Programmes makes its contribution independently of Parent Academic Environment in this model, while in the Advantaged model the two Parent predictors are closely correlated, adds support to the above interpretation of a relatively unintegrated set of skills in this disadvantaged sample.

Models 5 and 6

Disadvantaged Maths Groups ( 32 )

predicting

Maths Numeracy Attainment (Post-Test)

and

Maths Concepts Attainment (Post-Test)

**Sample:** All the Nursery children of the 32 mothers who attended the different mathematics programmes at five disadvantaged schools within the same inner urban education authority. The programmes consisted of eight one-hour sessions held at fortnightly intervals over a period of four months. The parent attendance at programme meetings was recorded and used to create the Parent Programmes variable in the path model.

**Predicting:** Model 5 Maths Numeracy Attainment at the end of the study period, this variable consisting of the scores on the WPPSI Arithmetic Test (modified to give a new ceiling of five advanced items).

Model 6 Maths Concepts Attainment at the end of the study period, this variable being a composite of the Boehm Concepts items and the Piagetian Tests, using standardised scores. The weightings for each test are defined in sub-section 6.831.

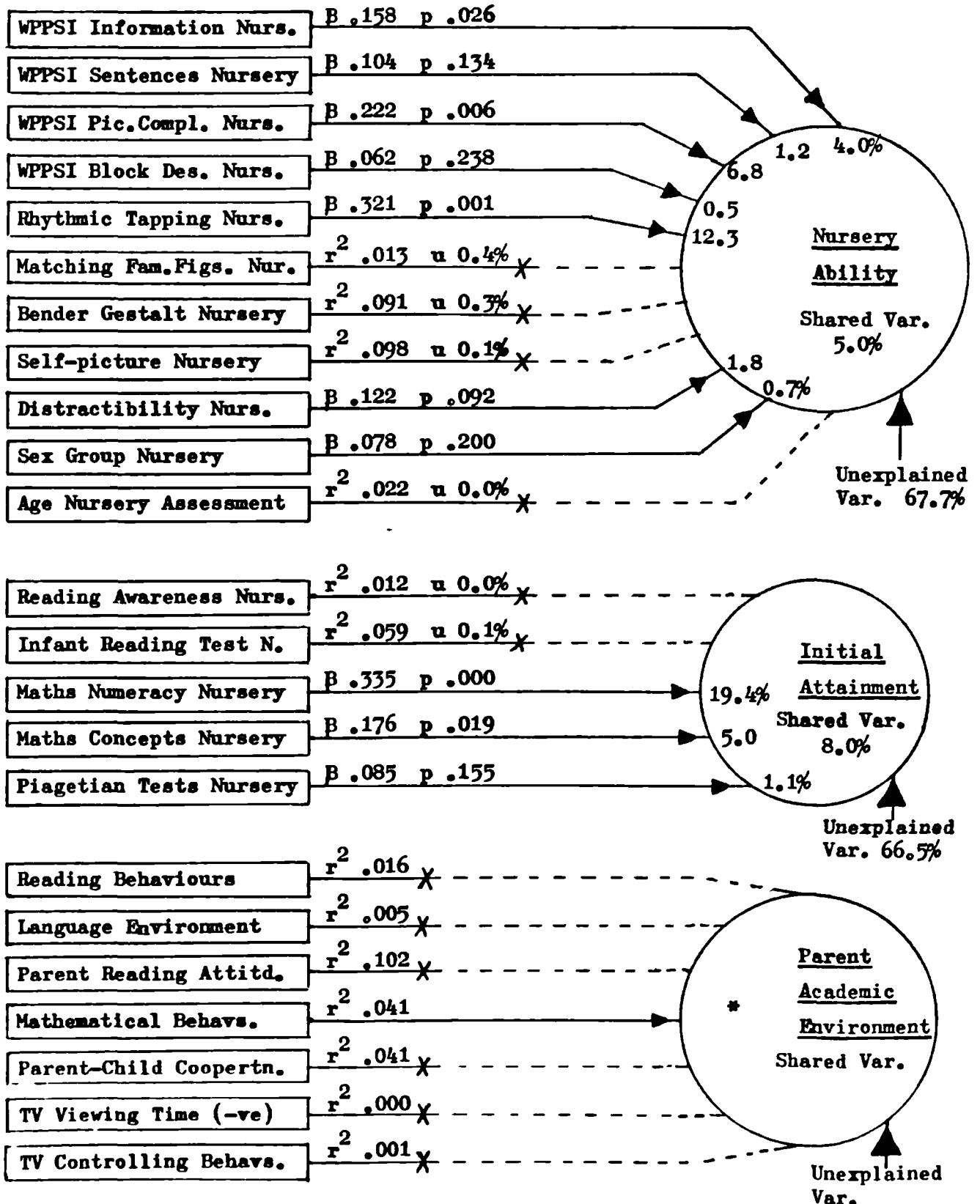
**Comparison of models:** In view of the fact that the same sample and the same independent variables are used for each model, it is possible to make direct comparisons between the models, identifying the differing contributions to variance within each model and studying the hypothesised reasons for these differences.

Creation of Latent Variables for Model 5

Sample: Disadvantaged Maths Programme Groups

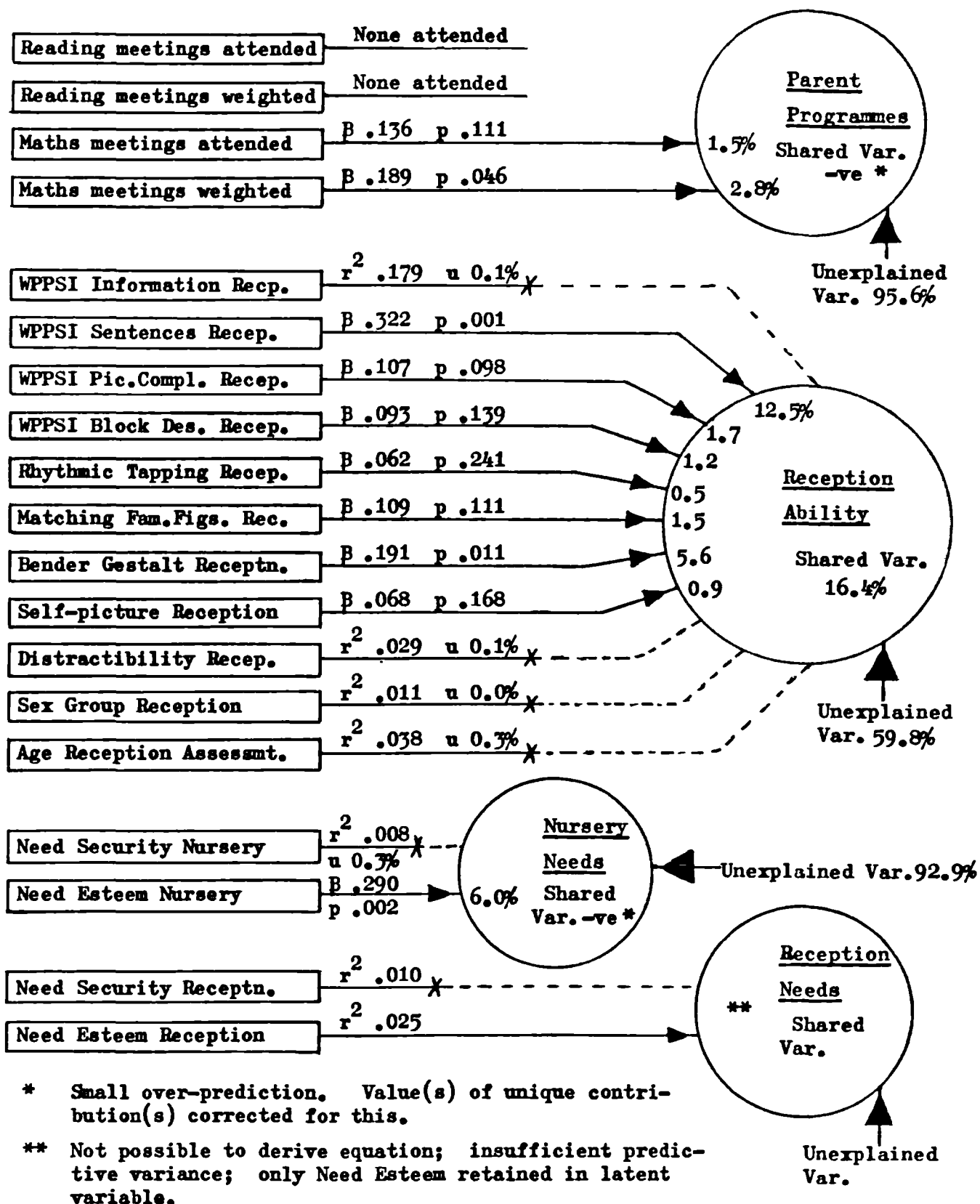
N = 32

Predicting: Maths Numeracy Attainment (Post-Test)



\* Not possible to derive equation; insufficient predictive variance; only Mathematical Behaviours retained in latent variable

Creation of Latent Variables for Model 5 (continued)



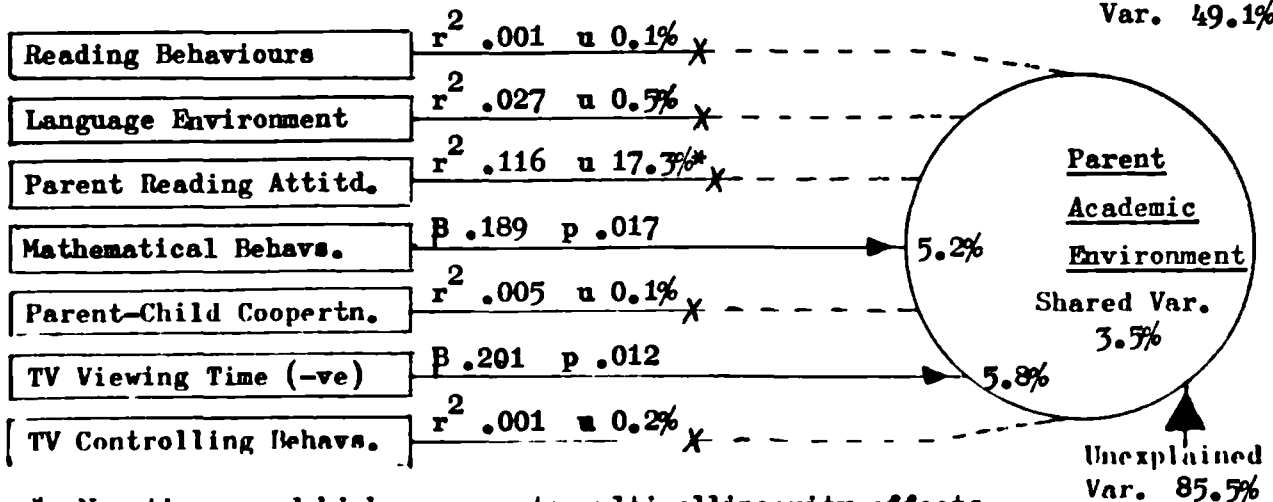
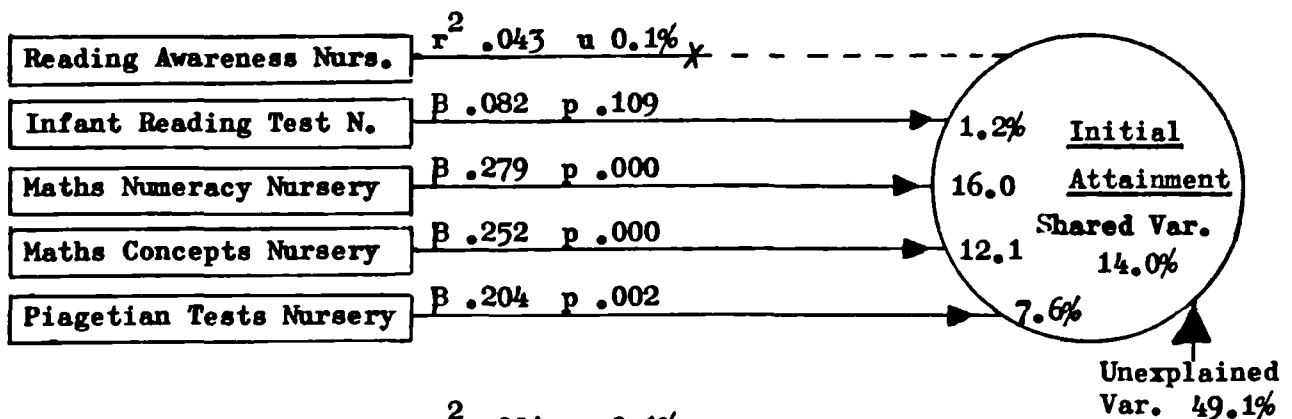
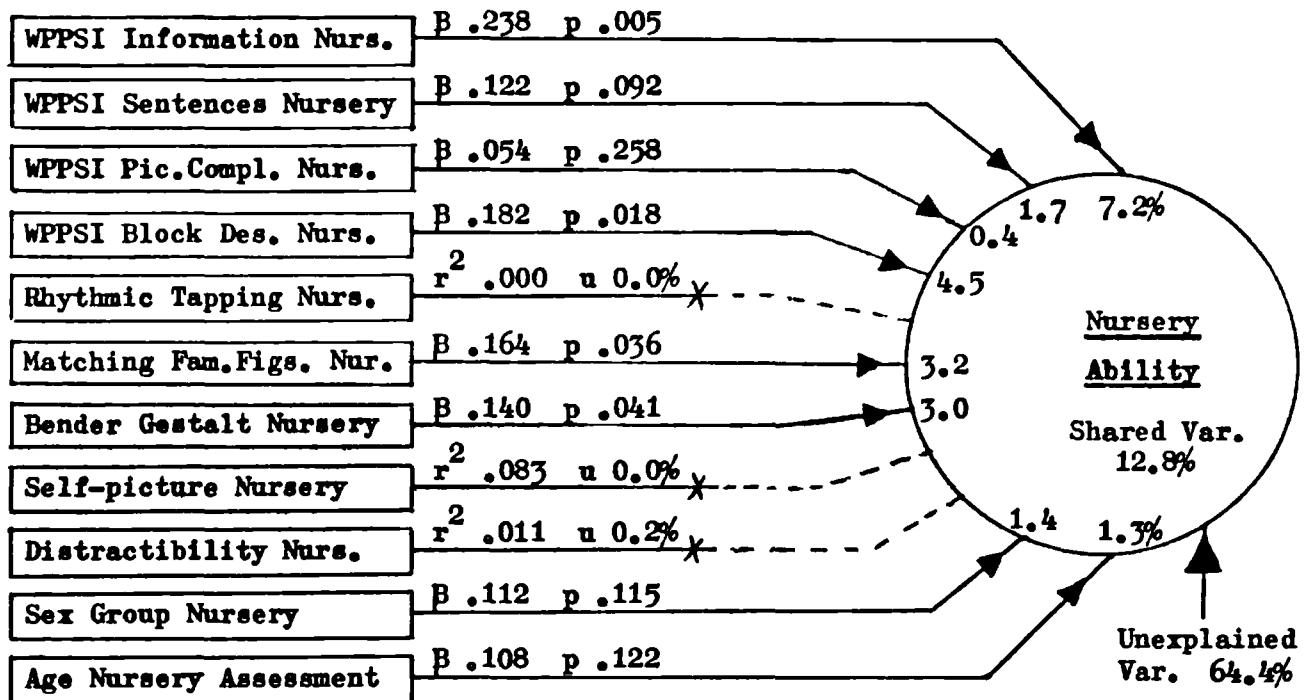


Creation of Latent Variables for Model 6

Sample: Disadvantaged Maths Programme Groups

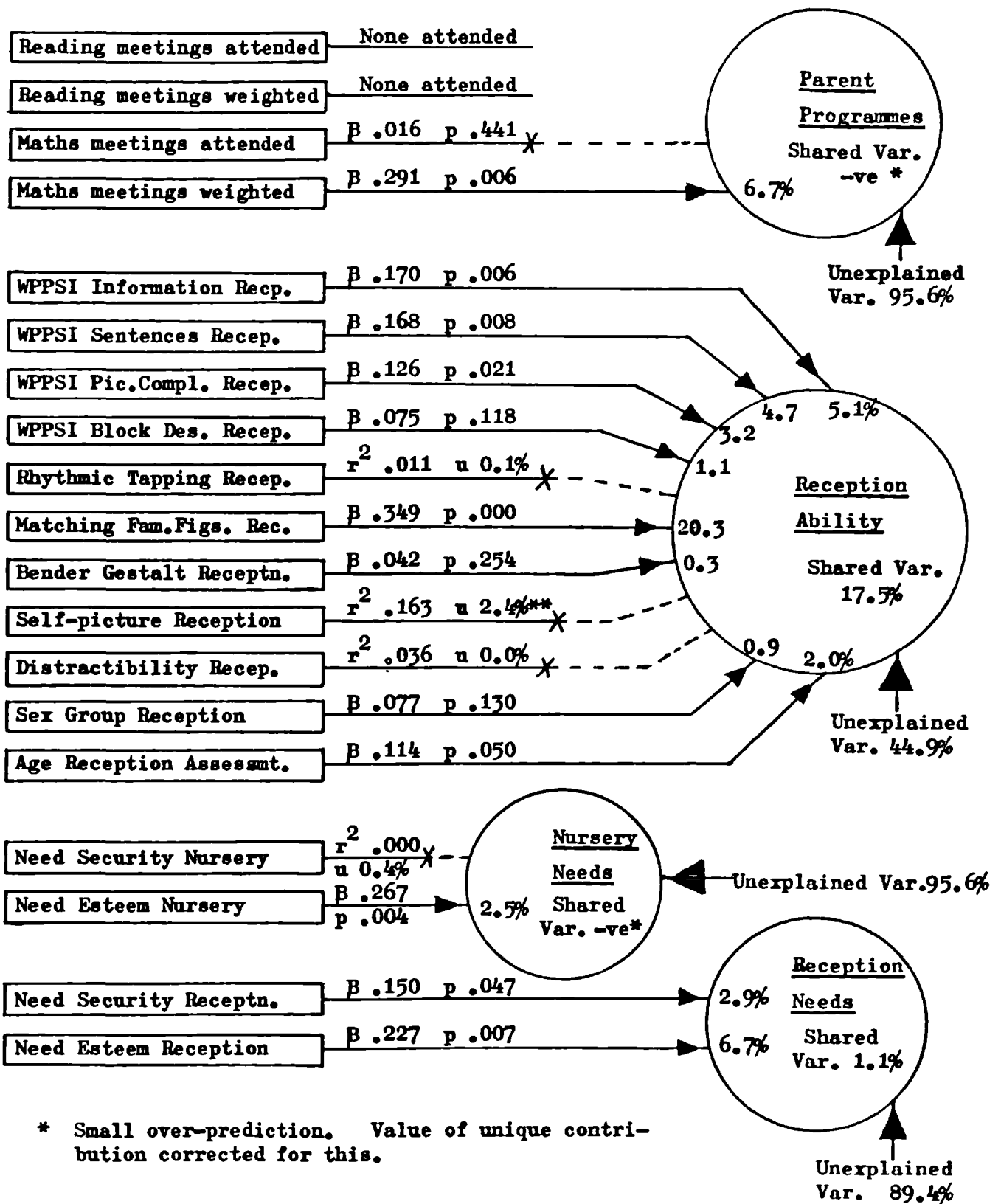
N = 32

Predicting: Maths Concepts Attainment (Post-Test)



\* Negative r and high u suggests multicollinearity effects

Creation of Latent Variables for Model 6 (continued)



Models 5 and 6: Creation of latent variables

The patterns of prediction for Nursery Ability and Reception Ability differ somewhat in the two models and each latent variable will therefore be handled separately.

In the Nursery Ability group of variables, WPPSI Picture Completion and Rhythmic Tapping are both strong predictors of Maths Numeracy but not of Maths Concepts, while WPPSI Information, WPPSI Block Design, Matching Familiar Figures and Bender Gestalt serve as strong predictors of Concepts but not of Numeracy. Age also appears a small predictor of Concepts, but not of Numeracy.

The picture changes somewhat for the Reception Ability group of predictors. Here the same set of tests was given to the same sample at a later point during the study (whenever a child entered the Reception class). For the prediction of Numeracy it is WPPSI Sentences and Bender Gestalt which are now the important predictors, while Rhythmic Tapping and Picture Completion are only minor predictors. It is understandable that rhythmic tapping, a skill which is thought to reflect sensitivity to musical rhythms, should be an early (Nursery) predictor of Maths Numeracy, but it is surprising that this variable should no longer appear as a sizable predictor in the Reception set.

In contrast the prediction pattern for Concepts has remained fairly stable (over the Nursery and Reception Ability sets), with an identical pattern of included and excluded variables. The only real differences are that the total outcome variance explained by this set has risen from 35 to 55 per cent, with M.F.F. now contributing a unique 20 per cent within this multiple regression equation; Bender Gestalt on the other hand has dropped to almost zero prediction.

The changes within the Concepts variable sets, between Nursery and Reception, are generally limited and could be expected to occur in a sample as small as this. However the changes in the pattern of predictors for Numeracy outcome are larger and suggest a certain instability in this particular model.

The size and nature of the contributors within the set of Initial Attainment variables are as could be expected from the definitions of the outcome variables. Maths Numeracy at pre-test level is the major contributor to Post-Test Numeracy (over 19 per cent) while Nursery level Maths Concepts (5 per cent) and Piagetian Tests (1 per cent) each make only small contributions. On the other hand with Maths Concepts as an outcome, the Nursery attainment variables of importance are Numeracy (16 per cent), Concepts (12 per cent), and Piagetian Tests (over 7 per cent), with Infant Reading Test being a minimal contributor. It can be noted that the Nursery measure of Numeracy contributes more to Post-Test Concepts than does the Nursery measure of Concepts itself. It can also be noted that the Initial Attainment prediction of Post-Test Numeracy explains some 33 per cent of

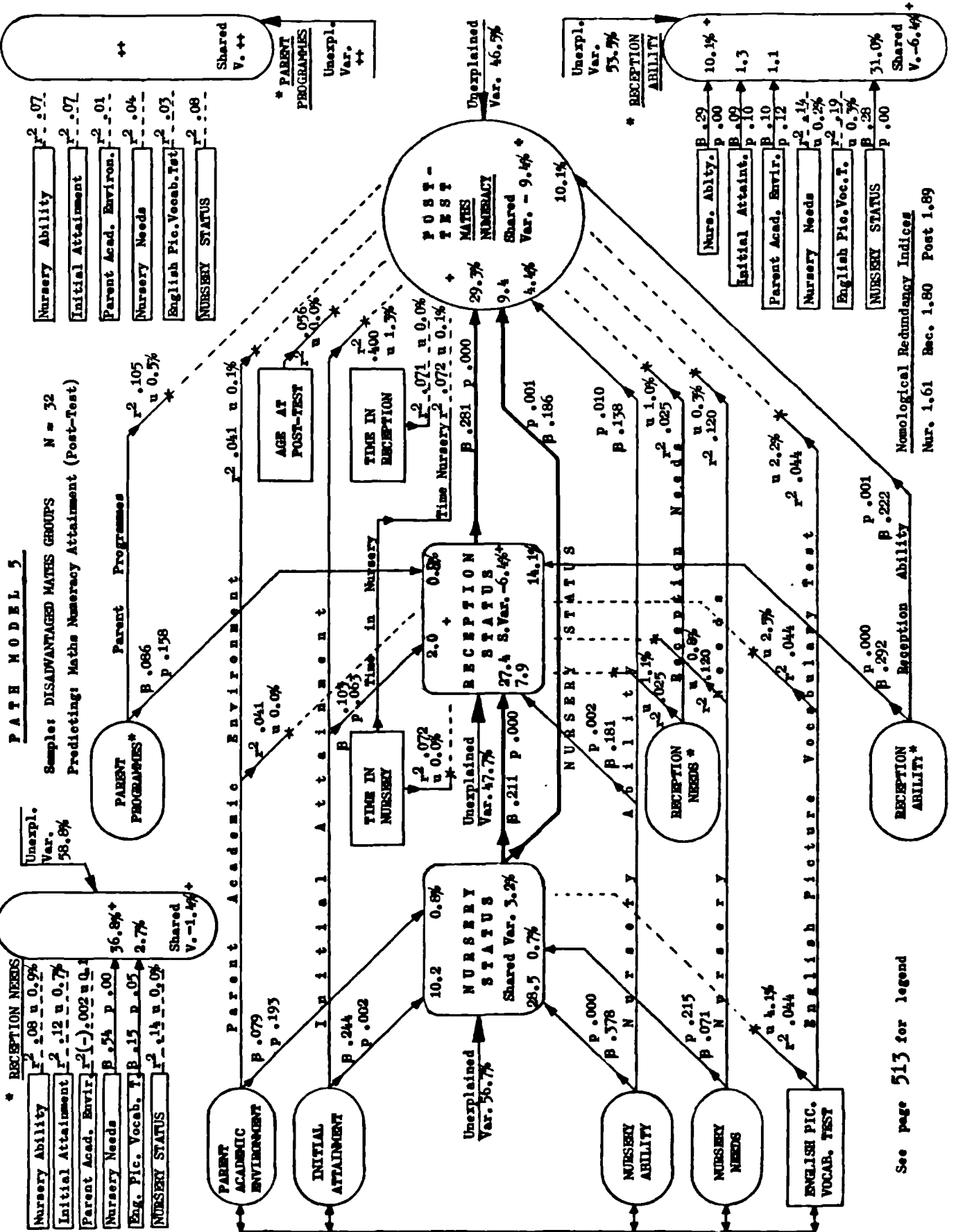
variance (i.e. in the absence of other competing predictors), while the equivalent Initial Attainment prediction of Post-Test Concepts explains 51 per cent of variance on its own.

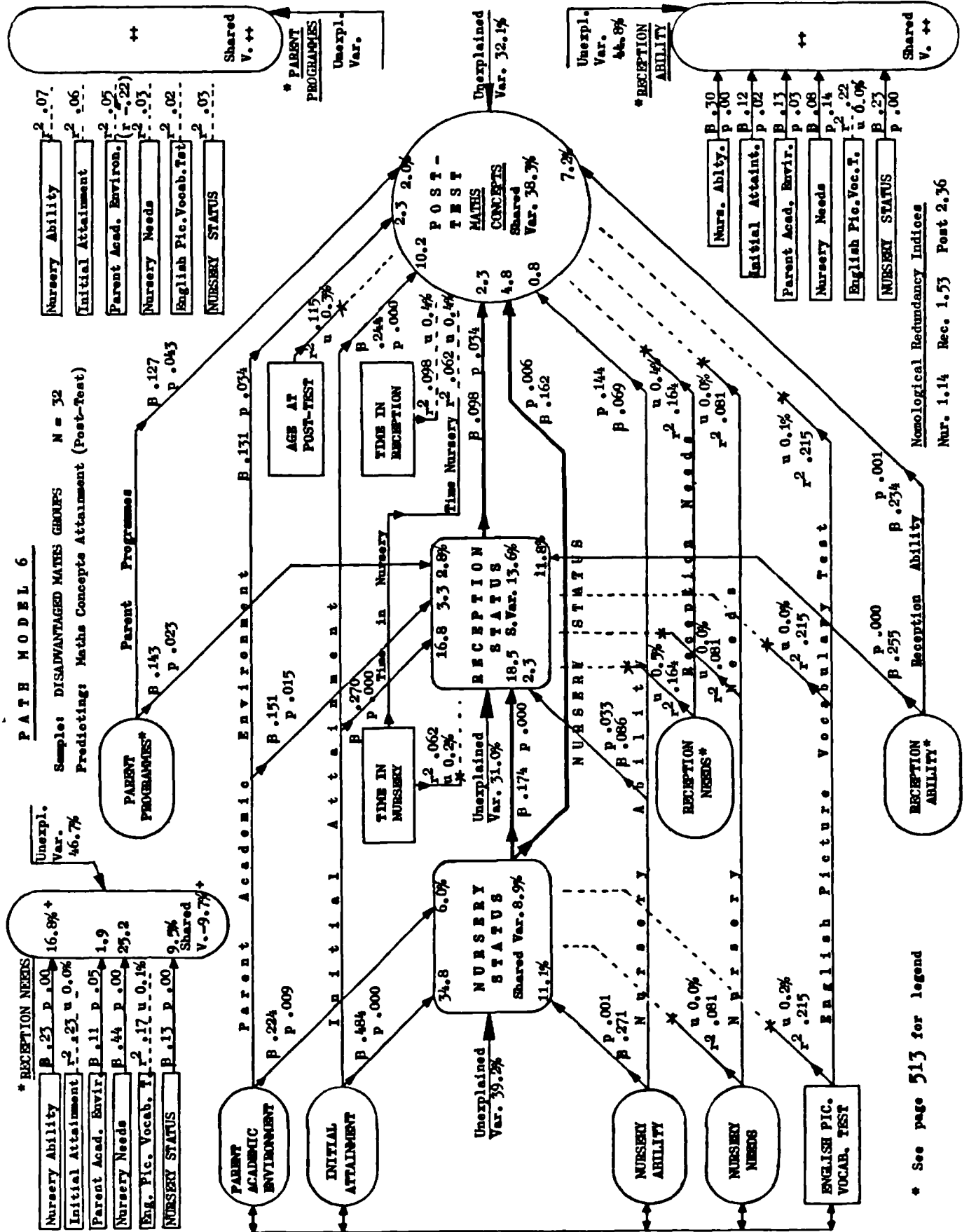
The set of Parent Academic Environment variables shows very considerable differences across the two models. Within the Numeracy model the prediction is close to zero, with the highest squared correlation of 0.10 coming from Parent Reading Attitude - something which has been seen throughout as a valueless predictor. (The simple correlation was in fact a negative one, suggesting that the obtained small relationship may well be due to chance.) With the low correlations in this set it was not possible to derive a regression equation and it was decided to retain only Mathematical Behaviours (squared correlation of 0.041 with outcome) as the only predictive variable within Parent Academic Environment in the Numeracy model.

The same latent variable performs quite differently when Concepts is the outcome. Here Mathematical Behaviours predicts 5 per cent of outcome and TV Viewing Time (scored negatively) nearly 6 per cent. This is the only model in which TV Viewing Time appears; it suggests, as one possible explanation, that children who view less television - for whatever reason - may have more opportunity for the kinds of manipulative games and activities which lead to improved performance on Piagetian Tests (one of the two component variables in Post-Test Concepts). However this hypothesis would need fuller investigation.

The Parent Programmes variables within each model show that Maths Meetings Weighted is the more important of the two meetings variables. The weighting, as explained earlier, was based on the amount of disturbance experienced (from infants present at the group meetings, or from outside interruptions) at each meeting. For both outcomes the meeting variables on their own predicted a little over 4 per cent of variance.

The Nursery Needs latent variable appears in both models, with Need for Esteem being the predictor of importance in each case. For the Concepts model the predictive power of the two Reception Needs variables in combination with each other is over 10 per cent, but again for the Numeracy model there is a rather surprising reversal in that neither of the Needs variables assessed at Reception level has any strong predictive power. Need for Esteem is retained to form the Reception Needs latent variable in this model.





Models 5 and 6: Construction of path models

The following points can be noted:

1. The total outcome variance predicted in the Maths Numeracy model (No. 5) rises from 43 to 53 per cent, from Nursery to Post-Test level; these figures suggest that there is a good deal of untapped variance in the model, possibly related to other predictors which have not been measured. The changing nature of some of the latent variables (e.g. Nursery and Reception Ability, and Nursery and Reception Needs) in the Numeracy model suggests a certain instability in the model, so that a good part of the unexplained variance may be simply error variance. Another unusual finding is that there are only four predictors in the final path regression in the Numeracy model. This is not linked to a high shared variance, since two of the three main path equations in this model show a small negative shared variance. The negative variance is itself unsatisfactory, though it would probably disappear with a larger N. (An alternative strategy for overcoming negative shared variance is the exclusion of one or more variables, for example those with probabilities above .05 at the Reception level. However the coefficients for each of the variables present in the final (Post-Test) equation have very low p figures, offering no real grounds for exclusion at that level.)

The position with the Maths Concepts model is more satisfactory, with predicted variance rising from 61 to 68 per cent; there is also an increasing percentage of shared variance across the three main path equations, suggesting a degree of stability of prediction of the outcome within this model.

2. For the Numeracy model the nomological redundancy indices are reasonably satisfactory, despite the relatively poor performance of the model as a whole. In retrospect it may be argued that this newly developed index should contain a 'penalising' factor for equations which yield negative shared variance despite the reasonable parameters of the variables retained in those equations. The indices for the Concepts model are also satisfactory, although the index for the final equation is a little above 2.

3. The performance of the Ability and Attainment latent variables differ very greatly across the two models. This is a most interesting finding. In the Maths Numeracy model the Nursery and Reception Ability variables are far stronger predictors than is the Initial Attainment variable, and this contrast remains at each level. Initial Attainment contributes 10 per cent at Nursery level, 2 per cent at Reception level and nothing at Post-Test level; by comparison Nursery Ability contributes 28 per cent at Nursery level, followed by 8 and 4 per cent at successive levels; Reception Ability contributes 14 and 10 per cent at its two levels.

In sharp contrast the Maths Concepts model shows Initial Attainment contributing 35, 17 and 10 per cent respectively at the Nursery, Reception and Post-Test levels, with Nursery and Reception Status carrying only a limited amount of variance into the Post-Test equation. On the other hand Nursery Ability contributes only 11, 2 and 1 per cent at the three successive levels, while Reception Ability contributes 12 and 7 per cent at its two levels of prediction.

There is a certain agreement with expectation in these findings, although the degree of difference is rather surprising. The Numeracy criterion, which has been shown to rest heavily on Ability predictors, demands not only numerical awareness but also a measure of logical or cognitive manipulation. (For example, one of the WPPSI Arithmetic questions asks: 'Johnny had three marbles and lost one. How many did he have left?') On the other hand the Concepts criterion, resting strongly on the prediction of Initial Attainment, involves two forms of skill which may depend particularly on experience and maturation. The modified Boehm Concepts Test requires knowledge of terms such as 'least', 'longer than', 'nearest', 'rectangle' etc.; the Piagetian Tests, which have been fully described in Chapter 4 and Appendix A3, require an awareness of the meaning of the concepts used in the tests and sufficient maturity to apply that knowledge to the task situations presented to the child.

4. The Parent Academic Environment plays an unexpectedly important part in the Maths Concepts model, while it hardly features in the Numeracy model. In the latter there is a limited contribution of less than 1 per cent ( $p .19$ ) at the Nursery level, but nothing further. In contrast the parent environment latent variable contributes reasonably though modestly and with low probability levels ( $p .009$ ,  $.015$  and  $.034$ ) at each successive stage in the path model. The difference between the Numeracy and Concepts models in this regard is not due to any statistical effect; the squared correlation of the Parent Academic Environment with Maths Numeracy is only 0.04, compared to the figure of 0.20 for the squared correlation of parent environment with Maths Concepts.

It is interesting to note that, apart from the inclusion of TV Viewing Time (negatively scored) in the parent environment latent variable for the Concepts model, the actual parent Mathematical Behaviours score is a combination of a set of simple number-oriented behaviours, the most prominent of which was the reported 'counting the steps as we climb into the flats'. This particular activity cannot be dismissed as a trifling and rather meaningless ritual, since it involves not only learning the basic number sequence but also, more importantly, matching numbers to objects instead of simply calling out a rote number sequence on its own.

What is of more concern, within the context of the disadvantaged models as a whole, is why is there no predictive power for Parent Academic Environment



within the Total Disadvantaged Sample nor even within the Disadvantaged Reading Groups? In both these models (1 and 4) the squared correlation with the outcome is only 0.1. It may be the case that while disadvantaged parents are prepared to participate with their children in activities such as counting while climbing stairs and other active behaviours related to number and number concepts ('give Susie more', 'find the shoe that goes with this one', and 'put out four spoons'), such parents are less willing to settle down to the seemingly sedentary behaviour of reading to their children, particularly when faced as many poorer parents are with the continuous need to contain the activities of several children within a relatively limited home environment.

It is interesting that the clear relationship between the parents' Mathematical Behaviours and the Concepts outcome should not have any spin off on the Numeracy outcome. The reason may well lie in the strong relationship which has been identified between Ability and the Numeracy outcome, with Ability reflecting more of the child's basic cognitive skills than its recent learned attainments.

5. The English Picture Vocabulary Test was not expected to predict strongly to either of the two Maths outcomes. Its squared correlation with Numeracy is very low (0.04) and it has no predictive power in that model. In the Concepts model its squared correlation with outcome is 0.215, but again it has no meaningful contribution in the model. The higher relationship does however support the hypothesis that the Concepts measure is more of an attainment than an ability variable, since E.P.V.T. is likewise a partial reflection of language learning (though clearly it does not reflect the totality of that process).

6. A second most important finding of this study, equal to that described in Model 4, is that the Parent Programmes variable predicts modestly to the Maths Concepts outcome at both the Reception and Post-Test levels. The amount of variance predicted is 2.8 per cent at the Reception level and 2.0 per cent at the Post-Test level, with probabilities of .023 and .043 respectively. This is a very satisfactory finding, given the limited extent of the parent programmes (eight one-hour meetings over a period of four months). It is noteworthy that, although E was given the impression that many of the participants in the maths programme groups would rather have taken part in the reading programme groups, and although the fall off from the maths meetings was greater than that from the reading meetings, the predictive value of the maths programmes in relation to the Concepts outcome is slightly greater than the parallel contribution of reading meetings to reading outcome in Model 4.

In contrast, the parent programmes were ineffective in altering the Maths Numeracy outcome; there is a minimal contribution of 0.9 per cent ( $p .158$ ) at Reception level and no meaningful contribution beyond that point.

It is worth noting that E set out to provide a maths programme which con-

sisted of a mixture of concept learning and number play, with the hypothesis that the main success of this programme would lie in the development of early numerical skills in the young children, provided the mothers played the games which were issued and explained at the programme meetings. The results show that it was in the area of concept learning alone that the programme achieved meaningful results, whereas the effect of the programme on numerical skills was negligible. This finding is, of course, in line with educational theory that concept learning has to precede any facility with numbers. It is interesting to see how strongly the results of the study reinforce this theory.

7. With the exception of a minor contribution from Nursery Needs at the first level in the Numeracy model (with  $p$  at an unsatisfactory .22), neither Nursery Needs nor Reception Needs make any real contribution to either the Numeracy or Concepts models, at any level.

8. It is surprising that neither Time in Nursery nor Time in Reception make any contribution to outcome in either the Numeracy or Concepts models. The squared correlations with the outcomes, ranging from 0.06 to 0.10 in these models, are certainly not large. (On the other hand a squared correlation of 0.11 between Time in Nursery and Post-Test Reading in Model 4 did yield a prediction of 4.2 per cent at the Reception level.) Why neither Time variable should predict to a Maths outcome is worth discussing.

It may be argued that the focus of both the Nursery and Reception curricula in the five disadvantaged study schools, and indeed in most of the other schools within the Metropolitan Authority, is mainly on general development and the fostering of pre-reading skills (in the Nursery) or reading skills (in Reception). The Reception work which does focus on numeracy is of necessity limited to the teaching of numbers, sorting and other manipulative activities - both physical and those involving paper and pencil - which are considered pre-requisites before number skills as such can be developed. This is of course part of the wider educational ethos in the United Kingdom which fosters the humanities first and only then the sciences.

It nevertheless remains slightly puzzling that neither Time variable contributed any meaningful variance to the Concepts outcome, since there was clearly some conceptual experience being given in the Reception class (and in the Nursery to a lesser extent). Marginal evidence for the increasing emphasis on number concepts can be seen in the fact that whereas the squared correlation of Time in Nursery with the Concepts outcome is only 0.062, the relationship between Time in Reception and Concepts increases to 0.098. The unique variance predicted in this model at the Post-Test level was 0.4 per cent for each Time variable, but the probability levels were not low enough to justify inclusion at that stage.

9. Within the Numeracy model Age makes no contribution at any level, even with-

in the two Ability latent variables. This is interesting, emphasising even more strongly the link between Numeracy attainment and Ability as assessed by the variety of cognitive and meta-cognitive measures used in the study.

In the Concepts model the Age variables do contribute within the two Ability latent variables (though adding only 1.3 and 2.0 per cent to variance within the Nursery and Reception Ability variable groupings). However Age at Post-Test makes no further contribution to outcome variance.

### Subsidiary regressions

There are some marked differences between the subsidiary regressions in the Numeracy and Concepts models. None of the three regressions involve Numeracy or Concepts as outcomes and one could therefore expect a certain similarity in the results. However each of the latent variables has weightings based on the predictive powers of the raw variables in relation to the Numeracy and Concepts outcome variables, so that in fact these latent variables do differ somewhat across the two models. - having thus rather different conformations in multivariate space, with differing predictive powers in relation to the outcomes.

The results show that Reception Needs, as constructed for the Numeracy model, is dependent almost entirely on the predictive power of Nursery Needs, with a small contribution from E.P.V.T. Within the Concepts model the regression of Reception Needs on the Nursery variables shows that Nursery Needs is still the main predictor, but Nursery Ability, Parent Academic Environment and Nursery Status all contribute a fair amount of variance to the Concepts outcome.

The regression of Parent Programmes on the Nursery level predictors shows that in both models this variable is not dependent on any of the prior predictors. This finding is in agreement with the earlier evidence that the Parent Programmes variable for the Disadvantaged Reading Groups was not predictable from the Parent Academic Environment of that sample. Here too the Parent Academic Environment latent variable has a negligible relationship with Parent Programmes; in the case of the Numeracy model the squared correlation is 0.01, while in the Concepts model the squared correlation of 0.05 conceals a negative correlation of -0.22. This should be seen against the fact that Parent Academic Environment did contribute quite well to outcome within the Concepts model; clearly both the Parent variables are quite distinct and not dependent on each other.

The regressions of Reception Ability on the prior variables are fairly similar across the two models. The Concepts equation was not honed down to the essential predictors (for reasons given earlier), but the regression coefficients obtained there show the equation to be not too dissimilar from the Numeracy equation.

### Overall findings

The following are the key findings which can be drawn from the two models, subject to the usual caveats in regard to the sample, the data and the analytical algorithms used here.

a. The parent programmes were effective for the development of Maths Concepts in the Disadvantaged Maths Groups children, but not for the development of Maths Numeracy. While the contribution to the Concepts outcome is a rather modest one it remains important. The actual contribution, as measured by the percentage of unique variance predicted, is higher than that predicted by parent programmes within the Disadvantaged Reading Groups sample. However the reading programmes contribution was made in competition with Time in Nursery and Time in Reception, both of which figured prominently in the Disadvantaged Reading Groups model. Neither Time variable appeared as a predictor in the Maths Concepts model. One further point is that Parent Academic Environment makes no contribution to attendance at Parent Programmes and in fact the correlation between the two variables is slightly negative.

b. There are major differences in the predictors of Maths Numeracy and Maths Concepts, as revealed by path analysis models. For Numeracy the key predictors are the Ability measures, with Initial Attainment playing an important but relatively small part and other predictors almost no part. In contrast the key predictor of the Concepts outcome is Initial Attainment, with Ability playing a much smaller though important role and the Parent variables - Parent Academic Environment and Parent Programmes - making useful unique contributions to explained variance. These fundamental differences in the models do suggest a basic psychological difference between the prerequisite skills required for the acquisition of mathematical concepts (including spatial concepts and Piagetian skills) and those required for mathematical numeracy (including the ability to handle and conceptualise number operations).

c. The Parent Academic Environment does contribute a meaningful amount to variance in the Concepts model, indicating that the simple spatial and number games identified by disadvantaged parents in the initial interviews are in fact important for the early development of mathematical concepts in their children. In contrast the other path models, particularly that focused on reading (Model 4), offer no evidence that the identified reading-oriented activities of disadvantaged parents make any meaningful contribution to post-test reading attainment. It is hypothesised that the sedentary nature of the task of reading to a child is more difficult in a home facing poverty and other social disadvantage, whereas active games involving manipulation, spatial terminology (under, above, next to, etc.) and basic number awareness rituals such as counting steps as they are climbed each day, are more likely to be followed naturally by parents, particu-

larly in a cultural environment dominated by visual media such as television.

d. Neither Time in Nursery nor Time in Reception make any meaningful contribution to the outcome in either of the Maths models. This suggests that the main focus of Nursery and Reception class experience is on the development of early reading skills (the contribution of both the Time variables to the latter is strong within the disadvantaged samples). However there is limited correlational evidence to suggest that the greater emphasis on Maths Concepts in the Reception class is reflected in the stronger relationship shown between Time in Reception and the Concepts outcome than between Time in Nursery and the same outcome.

Models 7 and 8

Disadvantaged Girls ( 71 )

and

Disadvantaged Boys ( 58 )

S A T E L L I T E M O D E L S

predicting

Total Attainment (Post-Test)

**Samples:** All the girls (Model 7) and all the boys (Model 8) in the Nursery classes of five randomly selected schools in disadvantaged areas of an inner urban education authority. A fuller description of this sample is given in the introduction to Model 1.

**Predicting:** Total academic attainment at the end of the study period, as defined in the introduction to Model 1.

**Satellite Models:** The principles followed in developing these satellite models have been fully set out in section 6.825. Both the models are closely patterned on the design of the Base Model (the Total Disadvantaged Sample), although the regression parameters derived within each model are unique to that model.

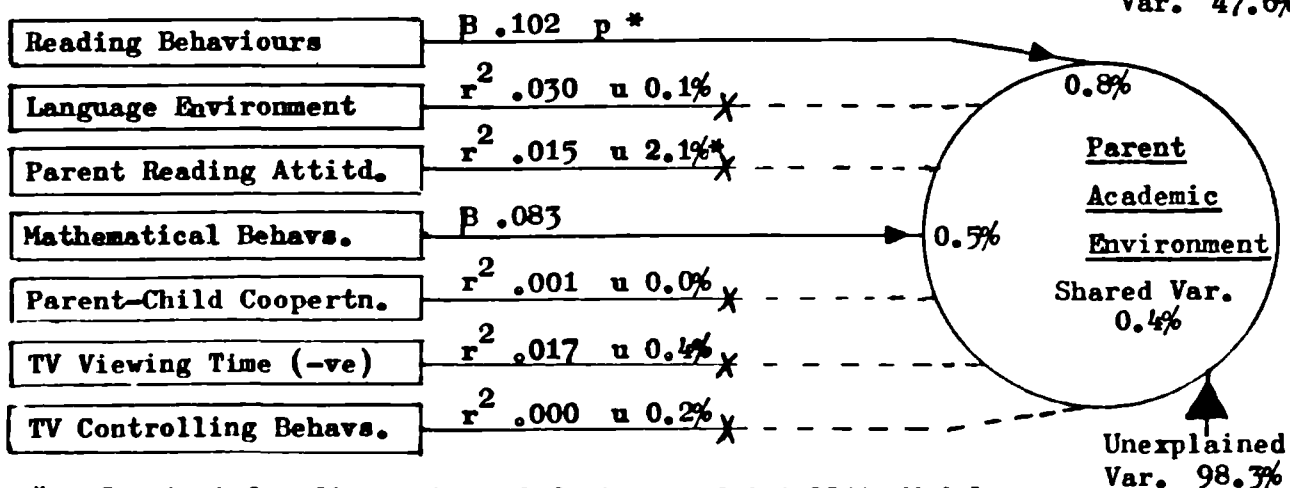
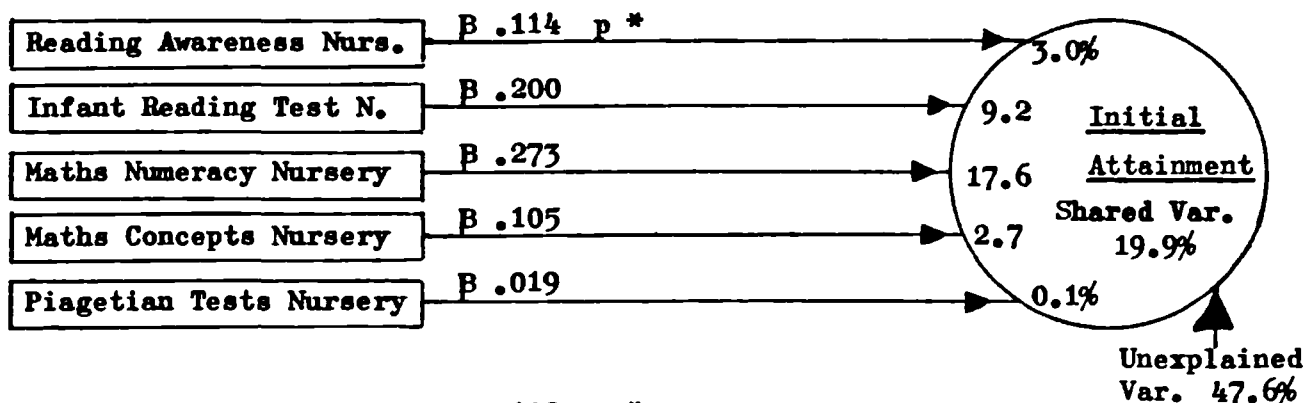
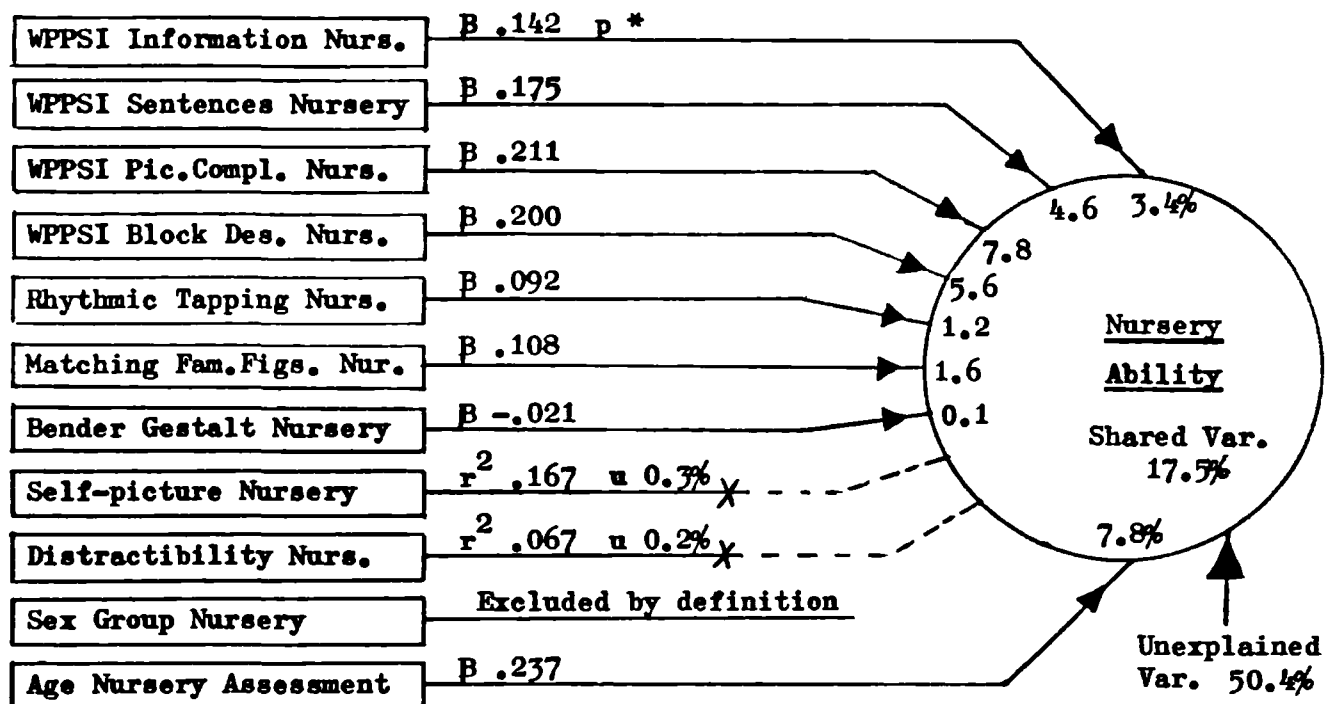
**Comparison of Models:** The particular value of the satellite modelling principle becomes apparent with the present comparison of the characteristics of the boy and girls samples within the Total Disadvantaged Sample. Since all the latent and status variables are constructed from sets of regression weights derived for the Base Model, and are therefore common to the Base Model and to all the satellite models, a direct comparison of each pair of regression equations (for the boy and girl samples respectively) can be based on these identically created sets of variables. The procedure simplifies the comparisons, although a more detailed study of boy and girl characteristics would demand independent models. Since there has already been a full discussion of the disadvantaged sample, this analysis will focus only on areas of principal difference between boys and girls.

Creation of Latent Variables for Model 7

Sample: Disadvantaged Girls (Satellite Model\*)

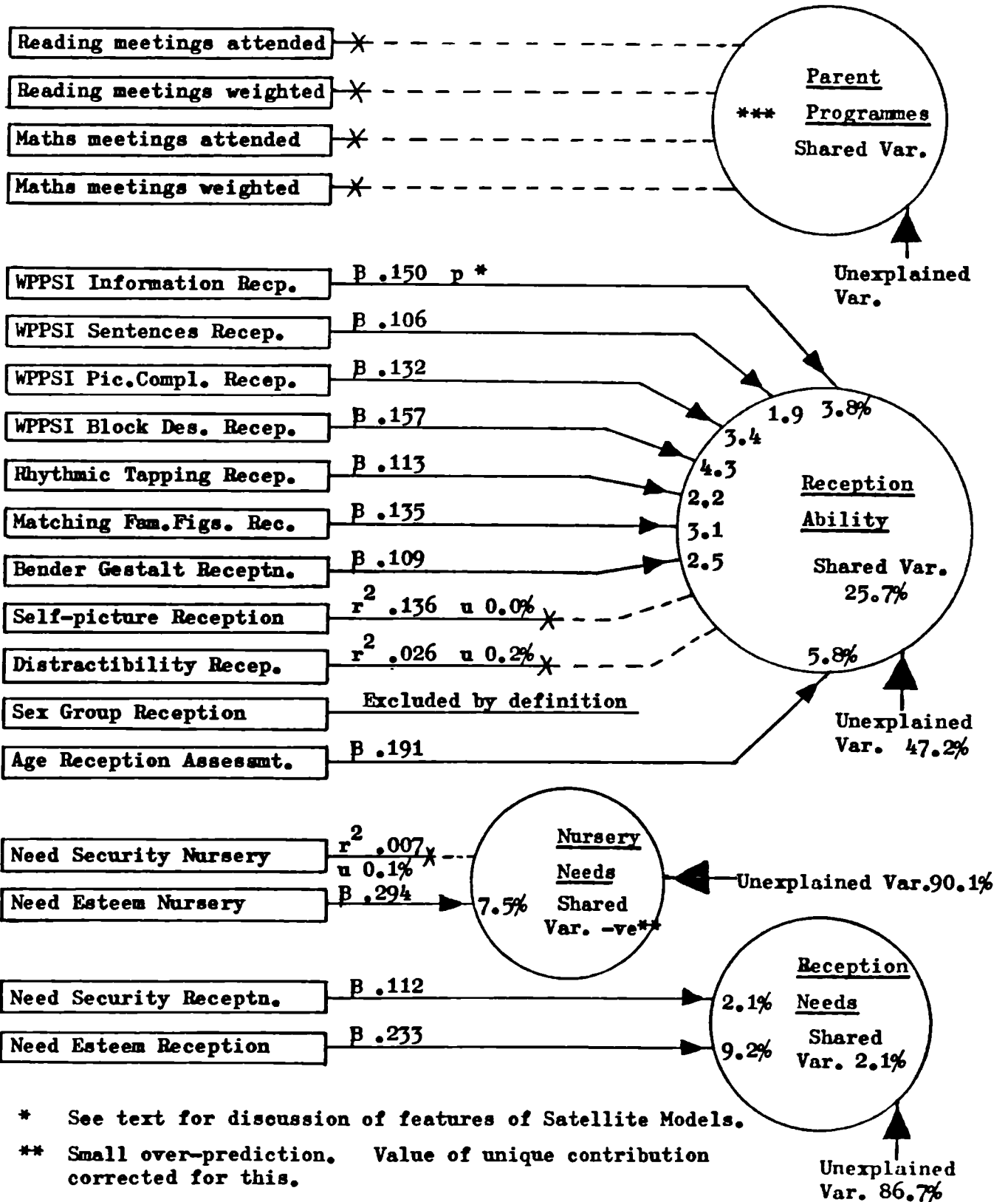
N = 71

Predicting: Total Attainment (Post-Test)



\* See text for discussion of features of Satellite Models

Creation of Latent Variables for Model 7 (continued)



\* See text for discussion of features of Satellite Models.

\*\* Small over-prediction. Value of unique contribution corrected for this.

\*\*\* Sample includes both attenders and non-attenders of Parent Programmes. Predictions too low to be of use.

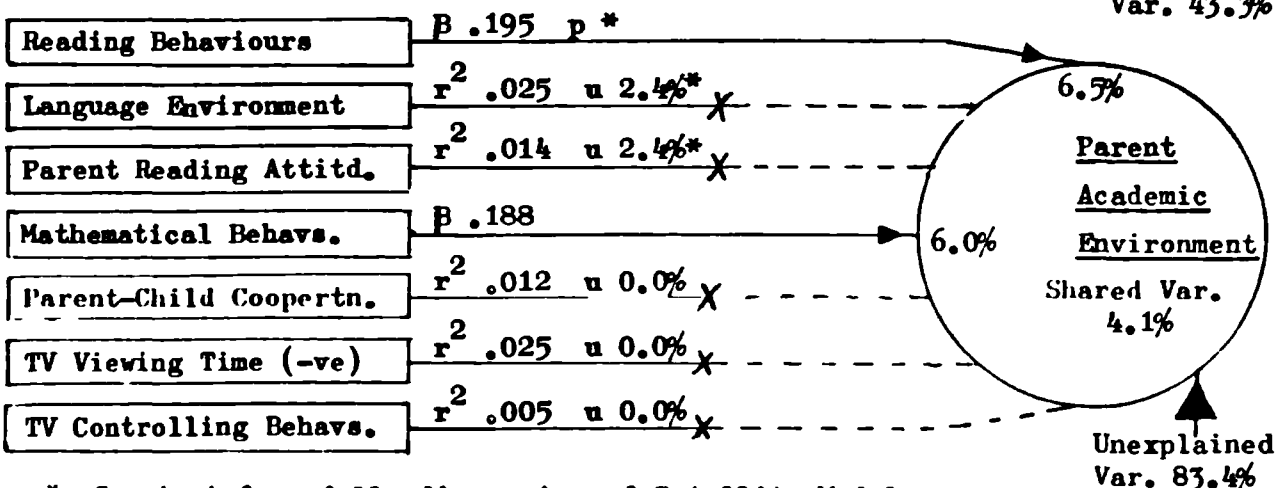
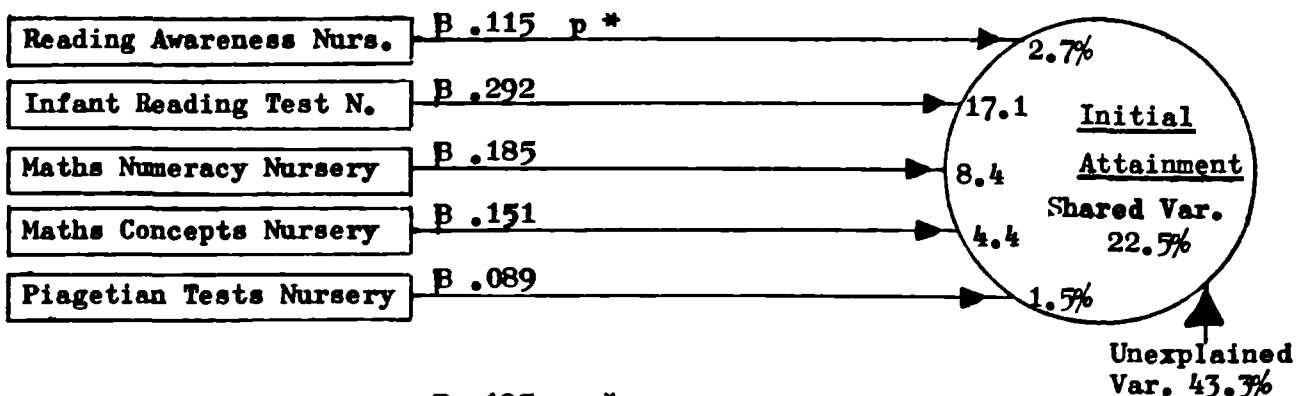
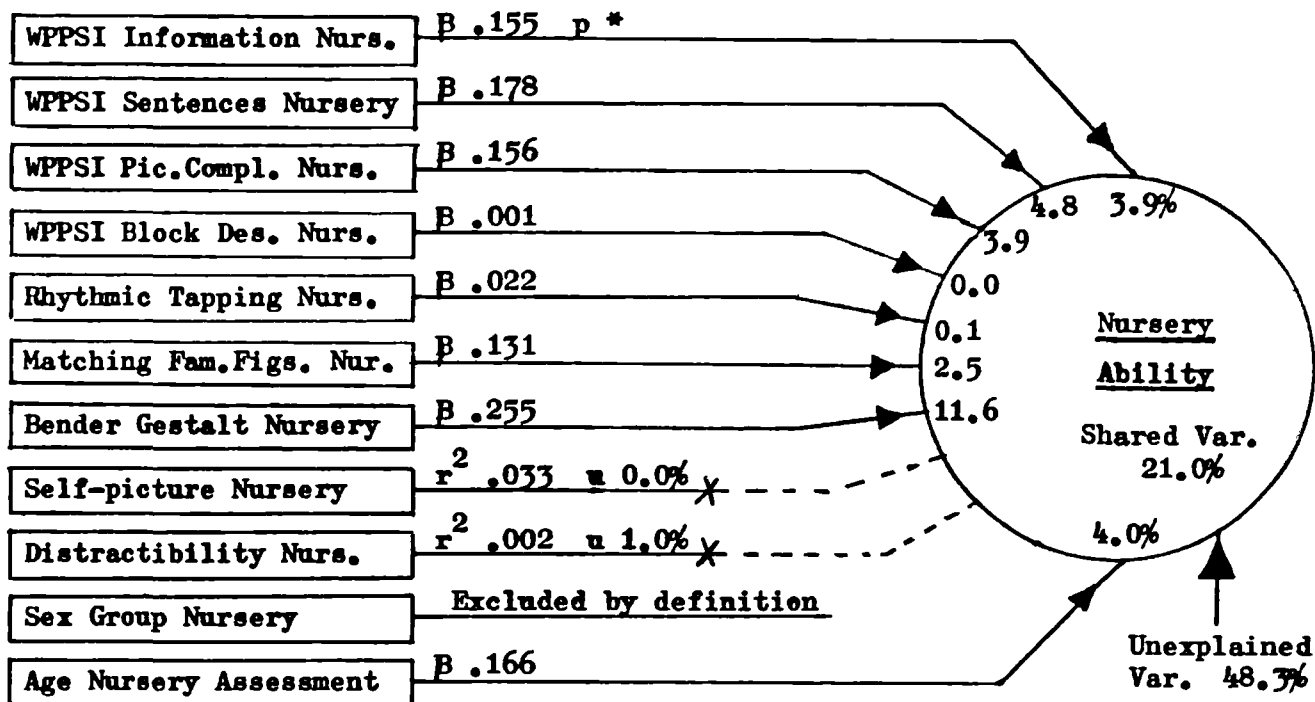


Creation of Latent Variables for Model 8

Sample: Disadvantaged Boys (Satellite Model\*)

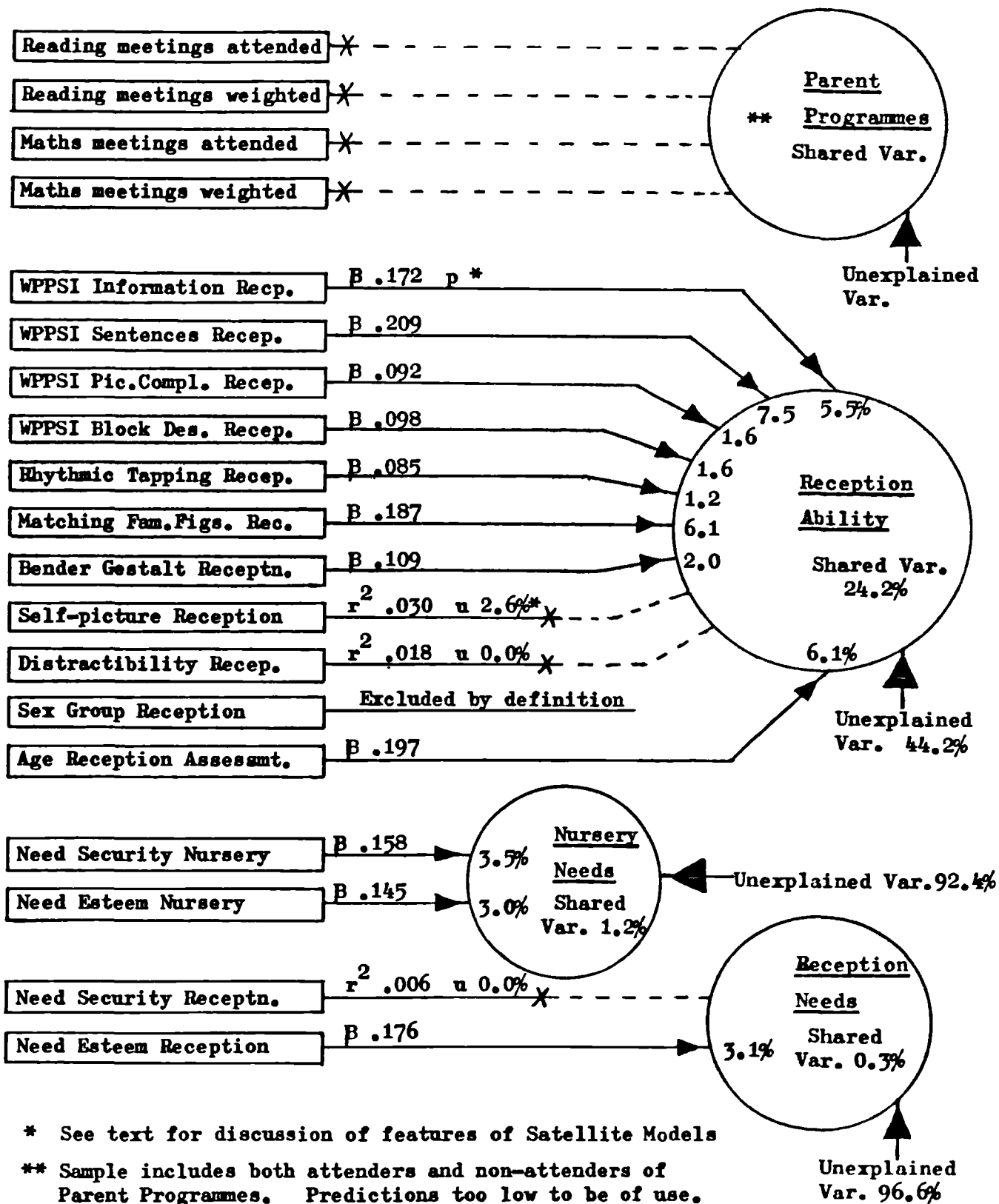
N = 58

Predicting: Total Attainment (Post-Test)



\* See text for full discussion of Satellite Models

Creation of Latent Variables for Model 8 (continued)



Models 7 and 8: Creation of latent variables

For the two Ability variables, at Nursery and Reception level, the pattern of the regressions is much the same as for the main Base Model sample. One difference occurs with the predictive power of the Bender Gestalt measure; at Nursery level the girls' performance on this has no predictive power for the post-test Total Attainment outcome, whereas for boys it contributes nearly 12 per cent of outcome variance within this group of predictors. Distractibility measured in the Nursery does offer a prediction of 1 per cent of variance for boys (although for the Disadvantaged Sample as a whole distractibility is not a meaningful predictor, the reasons for this having been discussed earlier). The age at which girls were assessed in the Nursery also appears to count more heavily than for boys. For the ability variable regressions at Reception level the predictive power of the Bender Gestalt no longer serves to distinguish the two samples, nor does age or distractibility. However WPPSI Sentences now appears as an important predictor for boys - over 7 per cent of outcome variance - but not for girls.

Regressions on the Initial Attainment variable set indicate that for girls Maths Numeracy is by far the most important academic predictor, but Piagetian Tests make no contribution to outcome variance, whereas for boys the Infant Reading Test is the strongest predictor. In other respects the regression parameters do not differ widely.

At this point it is useful to emphasise that big differences between the strengths of a variable's predictions are not usually due to either big differences in the mean levels of the variable across the two samples, nor to big differences in the two sample variances of the variables in question. Examination of a considerable number of interesting predictive differences has shown that the differences are usually less than half a standard deviation between the mean levels, with such differences not always in the expected directions. For example, a higher predictive power for a variable within one sample does not necessarily imply a higher mean level of that variable; the level may just as easily be lower than in the comparison sample. Given that there is no reason why a mean level as such should determine whether a sample variable is a good or poor predictor, the differences in the sample variances (or standard deviations) of a particular variable have also been examined for a number of variables. Here the differences across samples have been small and only occasionally of importance - where, for example, a sample with a bigger variance on some variable may be found to make a bigger contribution to an outcome. This is by no means a general rule, however.

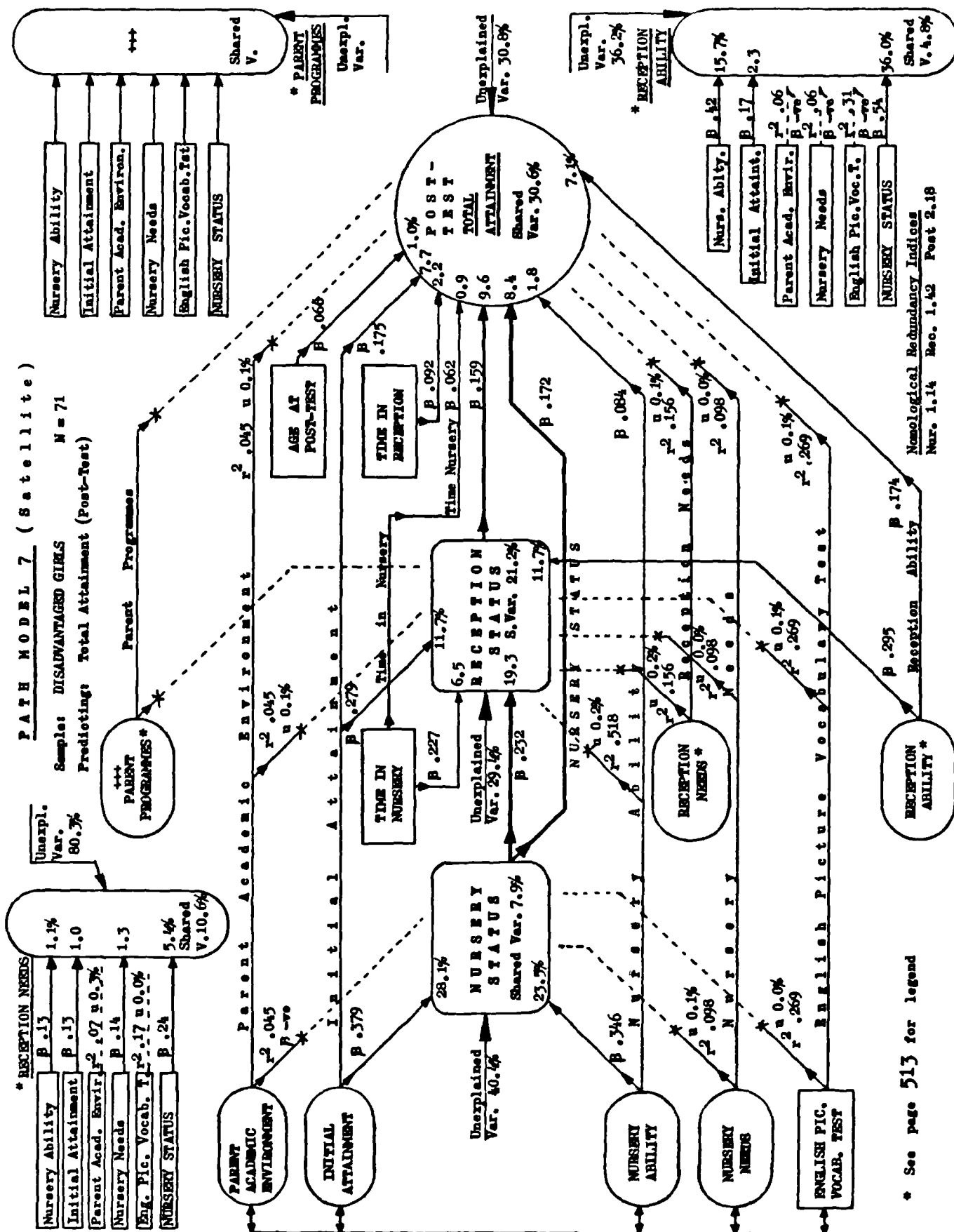
It has to be concluded therefore that differences in predictive variance are genuine model differences, based on differing patterns of relationships be-

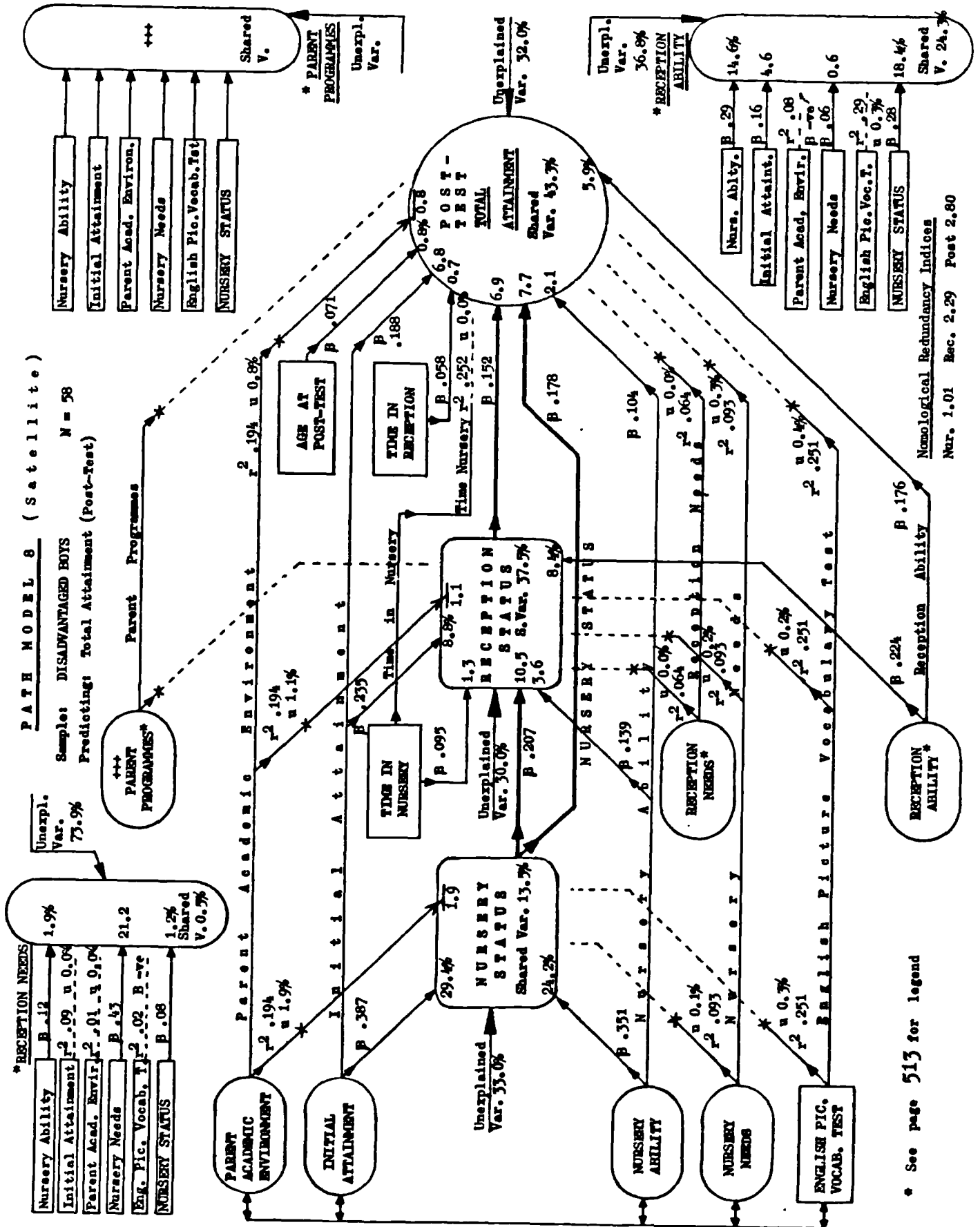
tween variables and within each model as a whole, rather than on different mean levels or variances of a particular variable. Indeed, attempts to find simple explanations in terms of powerful differences in a single variable are seldom valid. It is in the patterns of interrelationships, sensitively interpreted, that deeper understanding of differences is likely to be found.

A particularly strong difference between the two models lies in the regressions on the Parent Academic Environment variables. For girls none of the parent behaviours make any strong contribution to the outcome variance of Total Attainment at post-test. In total, Reading Behaviours and Mathematical Behaviours contribute only 1.7 per cent to outcome variance, in the absence of any outside predictors. For boys, in contrast, this set of variables predicts over 16 per cent of variance in the absence of other predictors, with Reading Behaviours and Mathematical Behaviours contributing equally. This is a potentially important finding and will be discussed later.

As explained earlier, in view of the composite nature of the Total Disadvantaged Sample, with differing parent programme groups and with two other groups who attended no programme meetings (Working Group parents and those who agreed to attend but failed to do so), the Parent Programmes variable makes no contribution to outcome variance within the sample as a whole. Consequently it was not taken into account in any of the sub-sample models which were divided by sex or ethnic group.

In the regressions of the outcome variable on to the Nursery Needs and Reception Needs sets, the predictive power of the girls' Needs measures was slightly higher than it was for boys at the Nursery level and considerably higher at the Reception level (13 per cent prediction of outcome variance for girls, in the absence of competing variables, compared to 3.4 per cent for boys). This suggests either that there is a greater teacher sensitivity to the emotional state of the girls, when scoring the Needs measures, or that the girls' needs, as interpreted by the teachers, are more closely correlated with their post-test performance than is the case with boys' needs.





\* See page 513 for legend

Nomological Redundancy Indices  
 Nur. 1.01 Rec. 2.29 Post 2.80

Models 7 and 8: Construction of path models

The following points can be noted, in comparing the regression parameters for the Disadvantaged Girls sample with those for the Disadvantaged Boys sample.

1. There is only a minimal difference in the amount of outcome variance predicted at each level in the two models.
2. Given that these are satellite models, the nomological redundancy indices are reasonably satisfactory in both models, with only the post-test index for the boys model being rather high (at 2.80). Since the models are limited to the design of the Base Model this is not a matter of serious concern.
3. There is almost no difference in the contributions of the Ability and Attainment latent variables across the two models, with Initial Attainment contributing a little more than Nursery Ability at the first level in both models - as it did in the Base Model - and continuing to make an important contribution at each subsequent level.
4. Despite the apparently high predictive power of Parent Academic Environment on its own in the boys' sample, when competing with other predictors in the path model this latent variable succeeds in making only limited contributions at each level of the model - ranging from 1 to 2 per cent of outcome variance. (It should be remembered that the Base Model does not incorporate any contributions from this latent variable in the two status variables, as Parent Academic Environment has no predictive power in that model.) In the girls' model the Parent Academic Environment makes no contribution whatever, the squared correlation with the outcome variable being only 0.045.

It is interesting to speculate why, with a Total Disadvantaged Sample which shows no measurable contribution of the parent academic environment to post-test Total Attainment, the boys' home environment should nevertheless make some measurable contribution to outcome variance, while the girls' environment does not do so. It is possible that even within a situation where there is but little awareness of the need to stimulate the children in academic or other terms, because of the pressure of poverty and other social disadvantage, boys may be the target for what little stimulation there is. This would be in accord with the long and questionable tradition in Western culture (and most other cultures) whereby the main educational and 'training' focus is centred on the boys, leaving the girls to develop their presumed role as carers of their siblings and as ultimate home-makers, with little awareness of the need for educational preparation for the latter profession.

5. E.P.V.T. continues to show a moderately high squared correlation with outcome in both models (between 0.25 and 0.27), but makes no meaningful contribution to variance. The possible reasons for this have been discussed earlier.

6. As already explained, Parent Programmes has not been retained as a variable in these or other satellite models.
7. Despite the moderately high predictive value of both the Needs variables on their own, when introduced into the path models in competition with other predictors they make no contribution to outcome variance at any level. It is therefore assumed that whatever shared variance there is between the outcome variable and the Needs variables is simply a reflection of other more meaningful predictive relationships such as those of Ability and Attainment.
8. A surprising finding in these models is that for girls Time in Nursery is an important predictor of outcome variance, with Time in Reception a moderately important predictor; for boys on the other hand the two Time variables make only a small contribution to variance. The 6.5 per cent contribution of Time in Nursery at Reception level, in the girls sample, is large enough to warrant a fuller discussion of the possible reasons for this.

It is the common experience of Nursery staff that boys are highly active in the Nursery environment and that most of the behavioural problems which arise originate from boys, so that much of the time and attention of teachers and Nursery Assistants is devoted to moderating these behaviour patterns. Although on the surface less staff attention may be devoted to girls, the attention which they do get may in general be focused more on their constructive activities such as examining picture books in the reading corner or manipulating blocks and other equipment intended to stimulate experimentation with and awareness of spatial dimensions. In E's experience it was far more often, though not always, boys who disrupted group activities and who on occasion needed to be taken away from the group by a Nursery Assistant for special caring.

In such a situation - a situation which demands immense dedication and commitment from a Nursery Staff who are dealing with pre-existing emotional and behavioural problems - the finding that Time in Nursery has a strong predictive value for girls but a much smaller one for boys, in terms of the academic outcome measure of Total Attainment, is in accord with E's experience of the Nursery situations.

9. The Age variables fulfil much the same predictive function at each point where they have been introduced and there is no evidence of a strong difference in predictive power across the two samples.

#### Subsidiary regressions

While the regression of Reception Ability on the Nursery level variables yields roughly comparable results for the two samples, the regression of Reception Needs on the same set of Nursery predictors shows that the level of Needs as assessed for boys in the Nursery is a very strong predictor of the level of



Needs as assessed in the Reception class; in contrast the prediction of Reception Needs for girls is based on nearly all the Nursery predictors, in particular Ability and Attainment.

This result suggests that for boys the pattern of Need for Esteem in particular remains identifiably the same in both Nursery and Reception classes, whereas the Needs as assessed in Reception for girls are very different from what have been assessed in the Nursery; this may call into question the validity of the Needs assessment for girls, while confirming the validity of the assessment for boys. There could of course be alternative explanations for the change in the Needs pattern, as seen by the Nursery staff who completed the Needs assessments. Alternatively the girls' Needs may well be related, both at Nursery and Reception levels, to their levels of Ability and Attainment (remembering that Needs are negatively scored, so that it is higher Ability and Attainment which predict to a higher Needs score - or, in absolute terms, to a lower level of the child's apparent need for esteem and insecurity).

#### Overall findings

The following are the key findings which can be drawn from the Disadvantaged Girls and Disadvantaged Boys models, subject to the usual caveats in regard to the sample, the data, the analytical algorithms used here, and the employment of satellite models rather than independent models.

a. The boys' Parent Academic Environment is a small but meaningful predictor of outcome variance, whereas the environment as assessed in the home interviews does not have any predictive power within the girls model. This suggests that in a situation of disadvantage, in which parent behaviours have little predictive value for post-test attainment, it is boys who benefit by the minimal stimulation which is present in the homes, rather than girls. There are social and historical reasons which may be adduced for this difference.

b. In contrast with the pattern in the home, Time in Nursery and Time in Reception make a much bigger contribution to final attainment for girls than they do for boys. This may be linked to the differing behaviour patterns of the sexes in the Nursery and Reception classes. Hutt (1974), for example, has shown that there are strong biological and other indications of the greater activity level of boys. In all the schools visited by E there was evidence of a higher level of manifest behaviour problems in boys from disadvantaged homes, as compared with girls. For Nursery and Reception staff this means that on average a higher proportion of their interaction time with boys is spent on social behaviour management, at the expense of the time that could otherwise be spent on fostering cognitive and academic development. Thus on average girls may well receive propor-

tionately more cognitively and academically oriented stimulation than boys.

c. In other respects the two models do not differ much and what differences there are are open to relatively simple interpretations.

Models 9 and 10

Disadvantaged Black Children ( 44 )

and

Disadvantaged White Children ( 62 )

S A T E L L I T E M O D E L S

predicting

Total Attainment (Post-Test)

**Samples:** All the children of West Indian and African ethnic origin (Model 9) and all the children of English, Irish and other European ethnic origin (Model 10) in the Nursery classes of five randomly selected schools in disadvantaged areas of an inner urban education authority. A fuller description of this sample is given in the introduction to Model 1.

**Predicting:** Total academic attainment at the end of the study period, as defined in the introduction to Model 1.

**Satellite Models:** The principles followed in developing these models have been fully set out in section 6.825. Both these models are closely patterned on the design of the Base Model (the Total Disadvantaged Sample), although the regression parameters derived within each model are unique to that model.

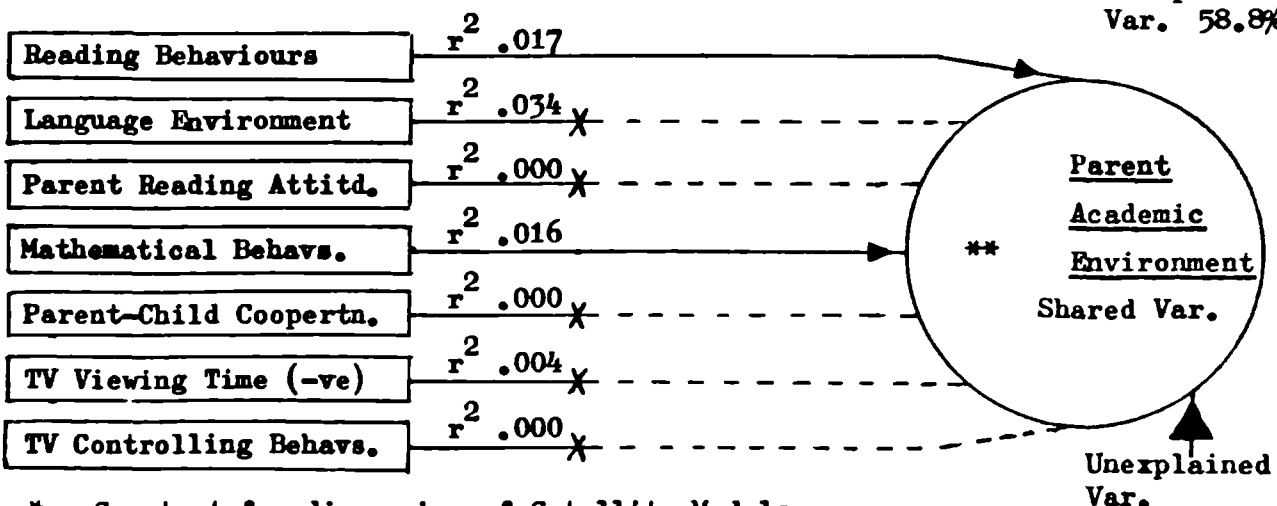
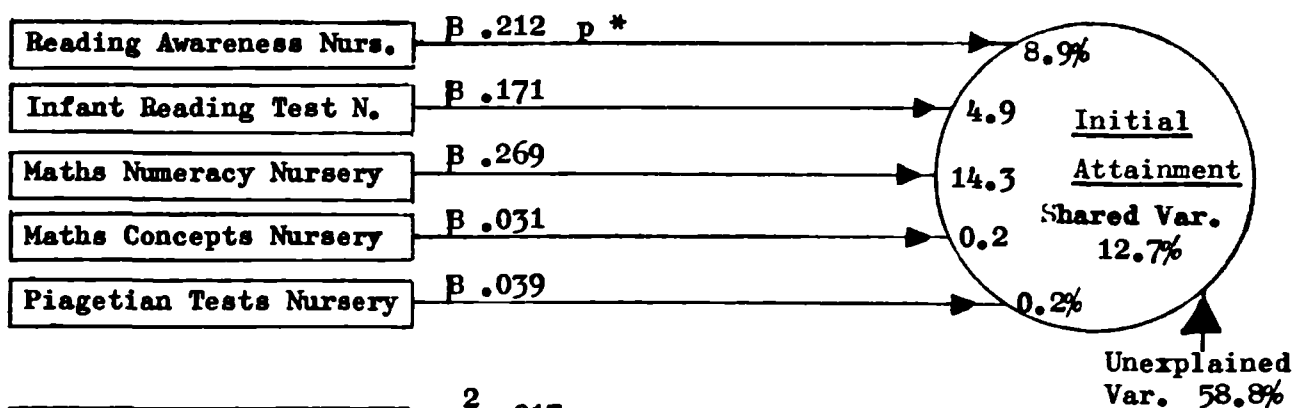
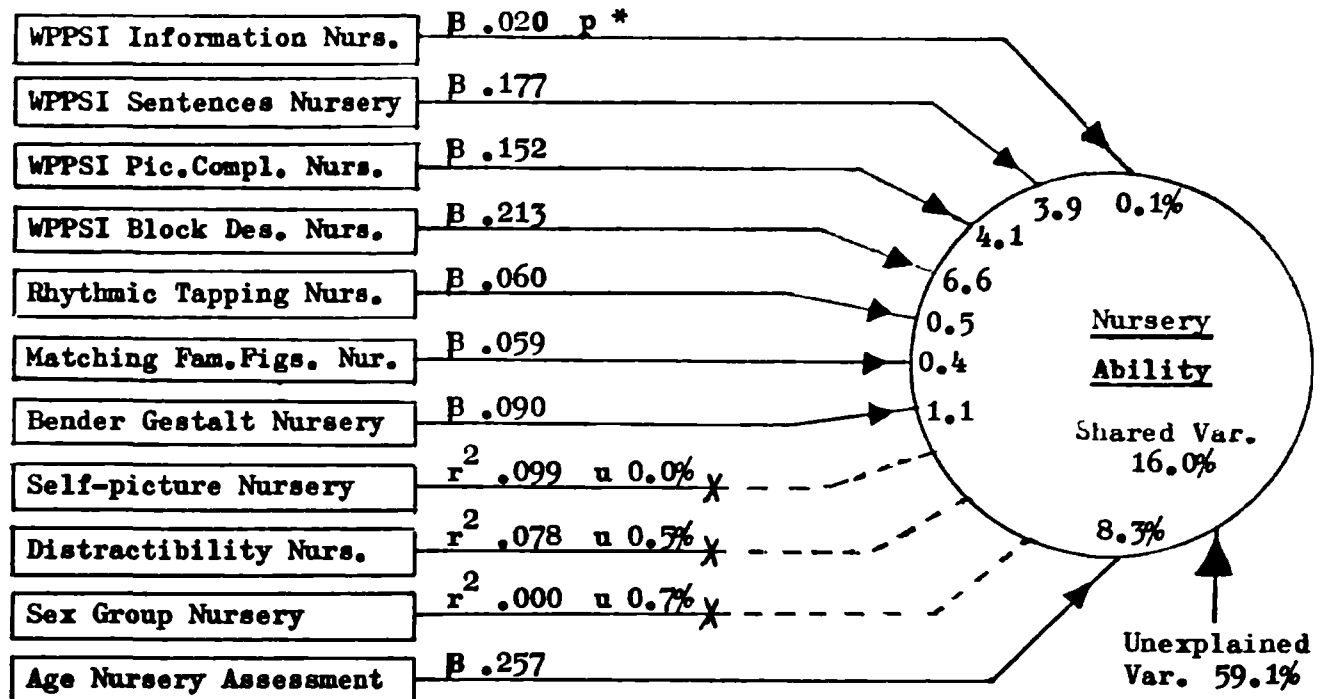
**Comparison of Models:** The rationale and value of comparing satellite models rather than fully independent models has been discussed briefly in the introduction to Models 7 and 8, and more fully earlier in this chapter.

Creation of Latent Variables for Model 9

Sample: Disadvantaged Black Children (Satellite Model\*)

N = 44

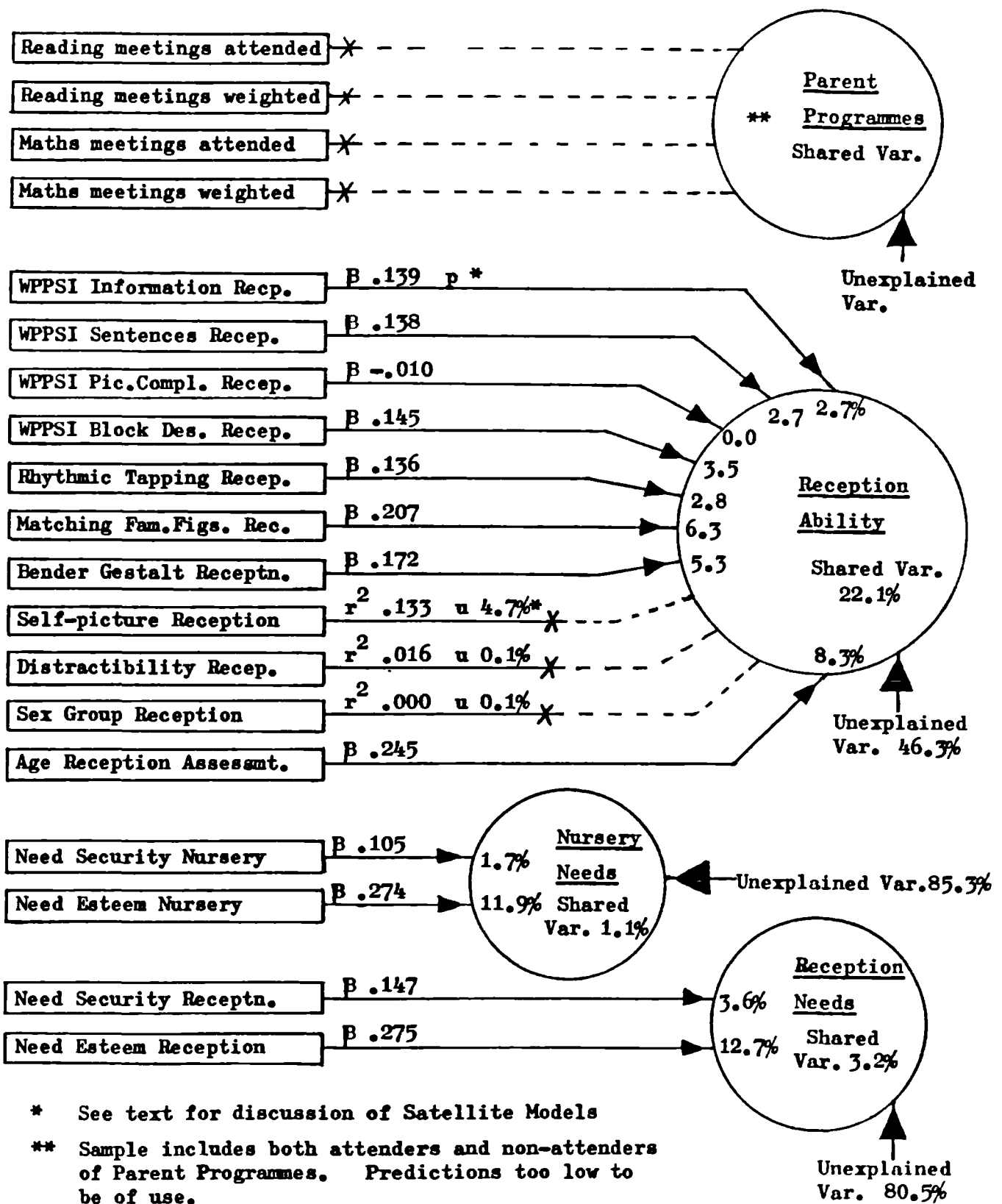
Predicting: Total Attainment (Post-Test)



\* See text for discussion of Satellite Models

\*\* Not possible to derive equation because of insufficient predictive variance

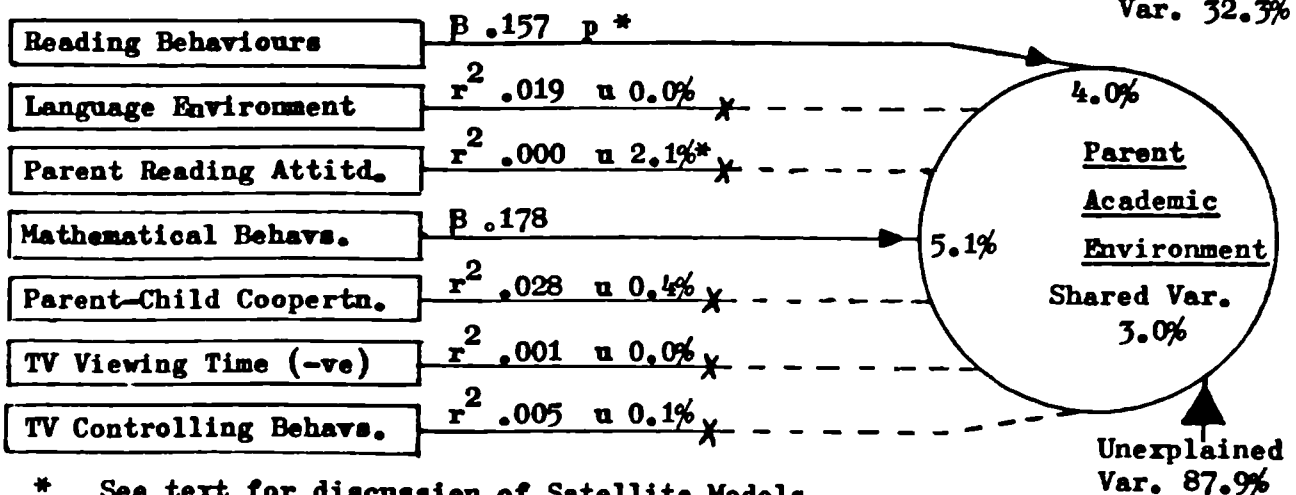
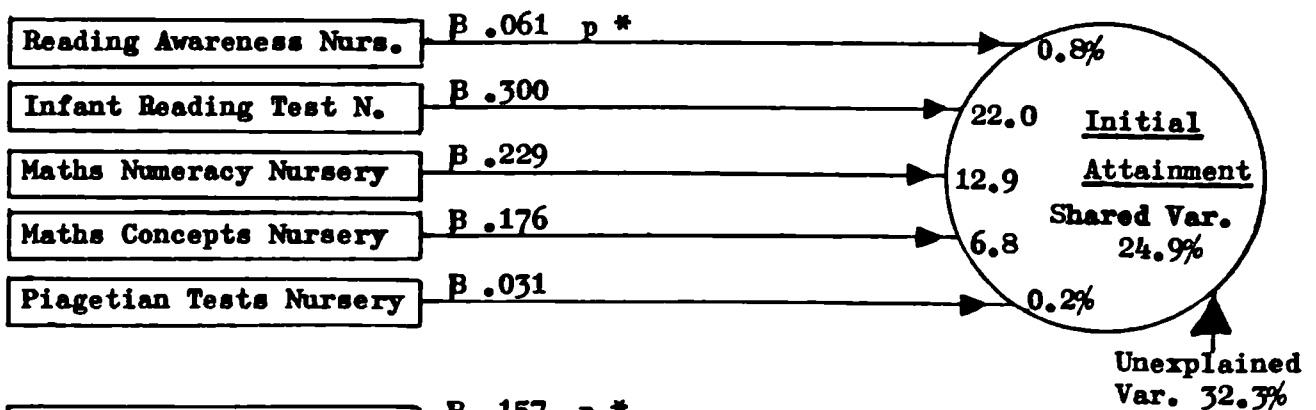
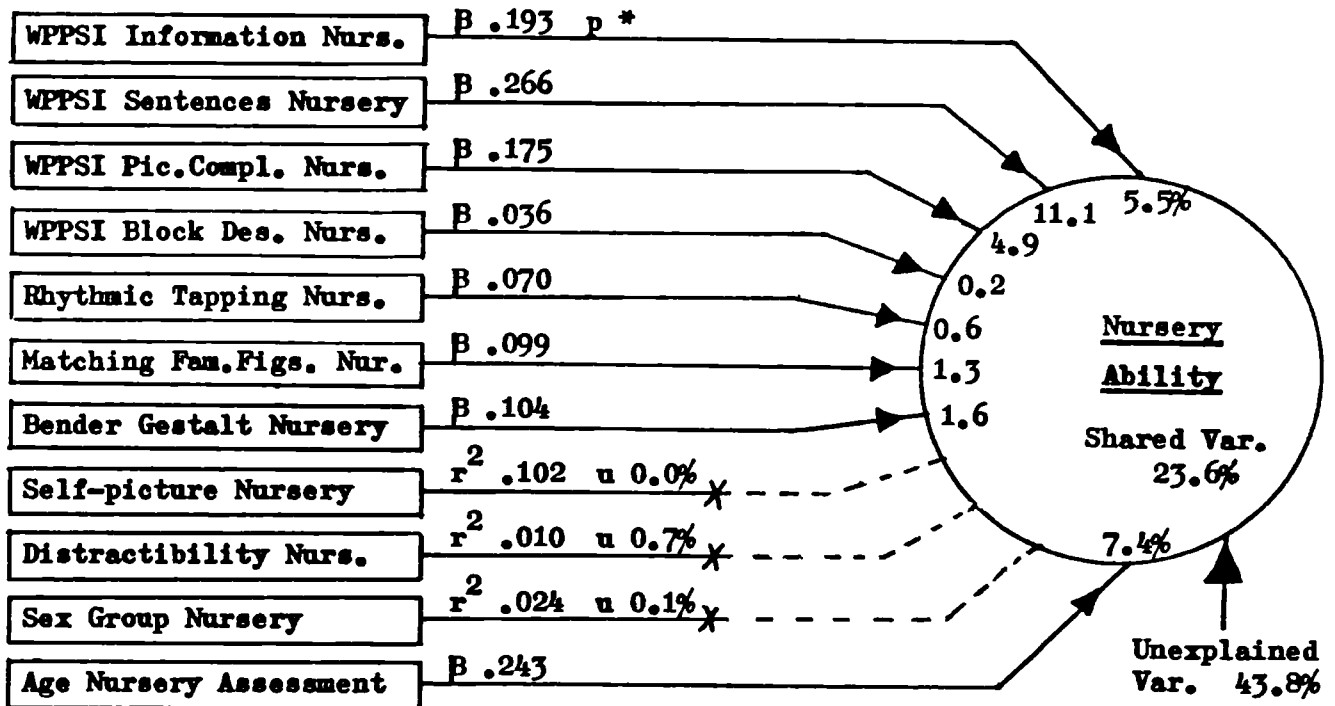
Creation of Latent Variables for Model 9 (continued)



Creation of Latent Variables for Model 10

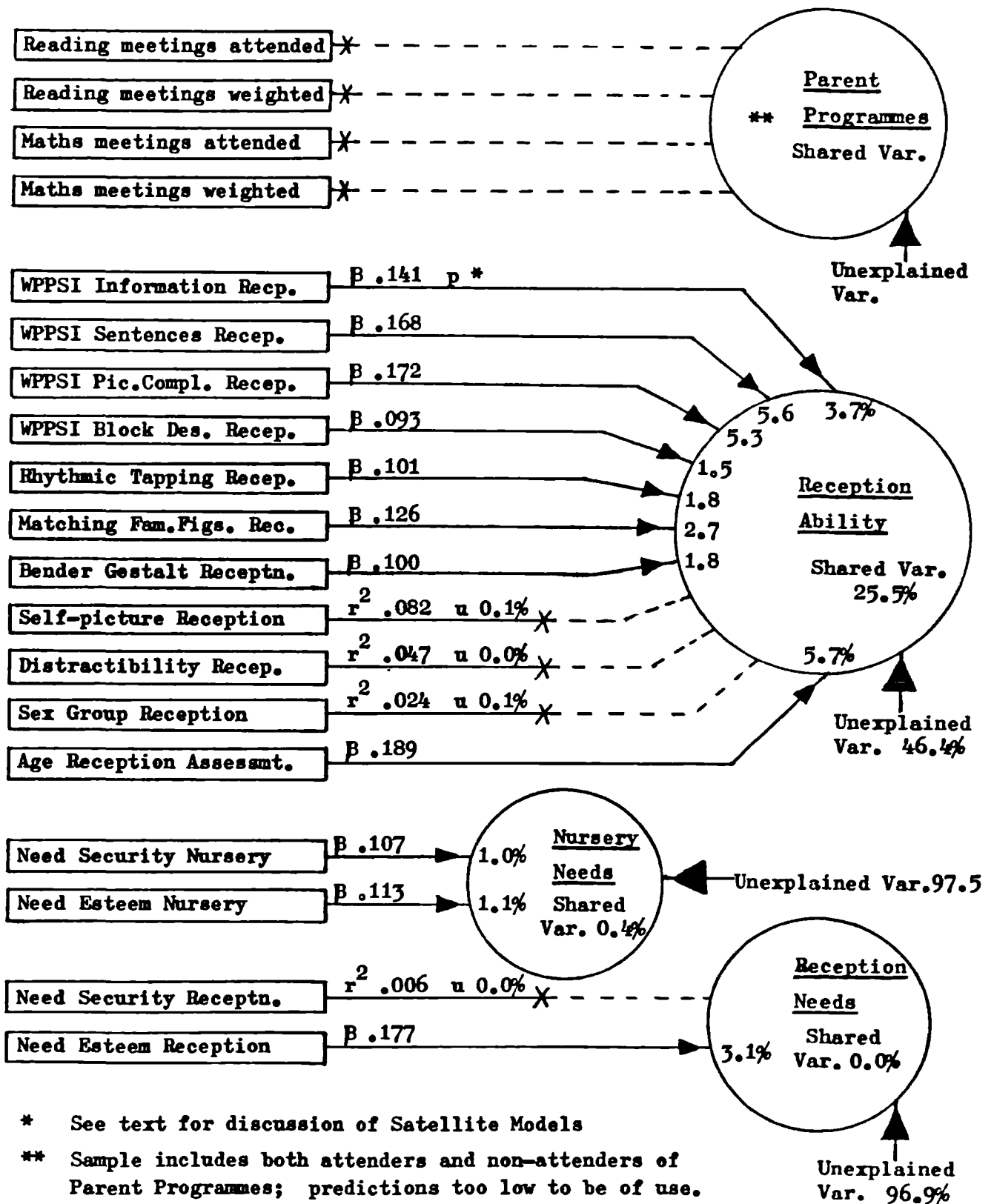
Sample: Disadvantaged White Children (Satellite Model\*) N = 62

Predicting: Total Attainment (Post-Test)



\* See text for discussion of Satellite Models

Creation of Latent Variables for Model 10 (continued)



### Models 9 and 10: Creation of Latent Variables

The regressions of the outcome variable on the sets of Ability variables for Black and White children show a fair measure of agreement, with only two persisting differences. WPPSI Sentences is a strong predictor for White but not Black children, while Block Design is a fairly strong predictor for Blacks but not for Whites. Apart from these differences the Nursery Ability set predicts some 41 per cent of outcome variance for the Blacks, compared with 56 per cent for the Whites. There are even fewer differences in the Reception level Ability variable set.

Differences between the samples in the regression equations using the raw Initial Attainment variables are more marked, with Infant Reading Test being a particularly strong predictor for White children, and Maths Concepts a smaller but also a strong predictor for this group; in both groups Maths Numeracy is a strong predictor, while for Black children Reading Awareness is a moderately strong predictor. The big difference here is in the total predictive strength of these equations - 41 per cent for Blacks and 68 per cent for Whites.

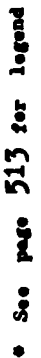
The biggest difference of all appears in the Parent Academic Environment regressions, where Reading Behaviours and Mathematical Behaviours each yield virtually no prediction for Black children, compared with the predictions of 4 per cent and 5 per cent respectively for the White children. The squared correlations between each of the two parent Behaviour variables and the outcome variable are 0.017 and 0.016 for Black children. As was pointed out in the previous comparison (Models 7 and 8), a large difference across samples in the predictive power (or even in the correlations with outcome) of a particular predictor does not necessarily indicate a difference in mean levels or in variances of the variable in question. But it may well indicate that the kinds of parent behaviours reported in the interviews were not predictive of early school performance. For example, parents in one ethnic group may tend to emphasise certain social behaviours, such as caring for younger siblings and other desirable non-academic behaviours in the home, at the expense of emphasising what have been termed Parent Reading and Mathematical Behaviours.

The reasons for the omission of the Parent Programmes variable from these satellite models have been explained in the discussion of models 7 and 8.

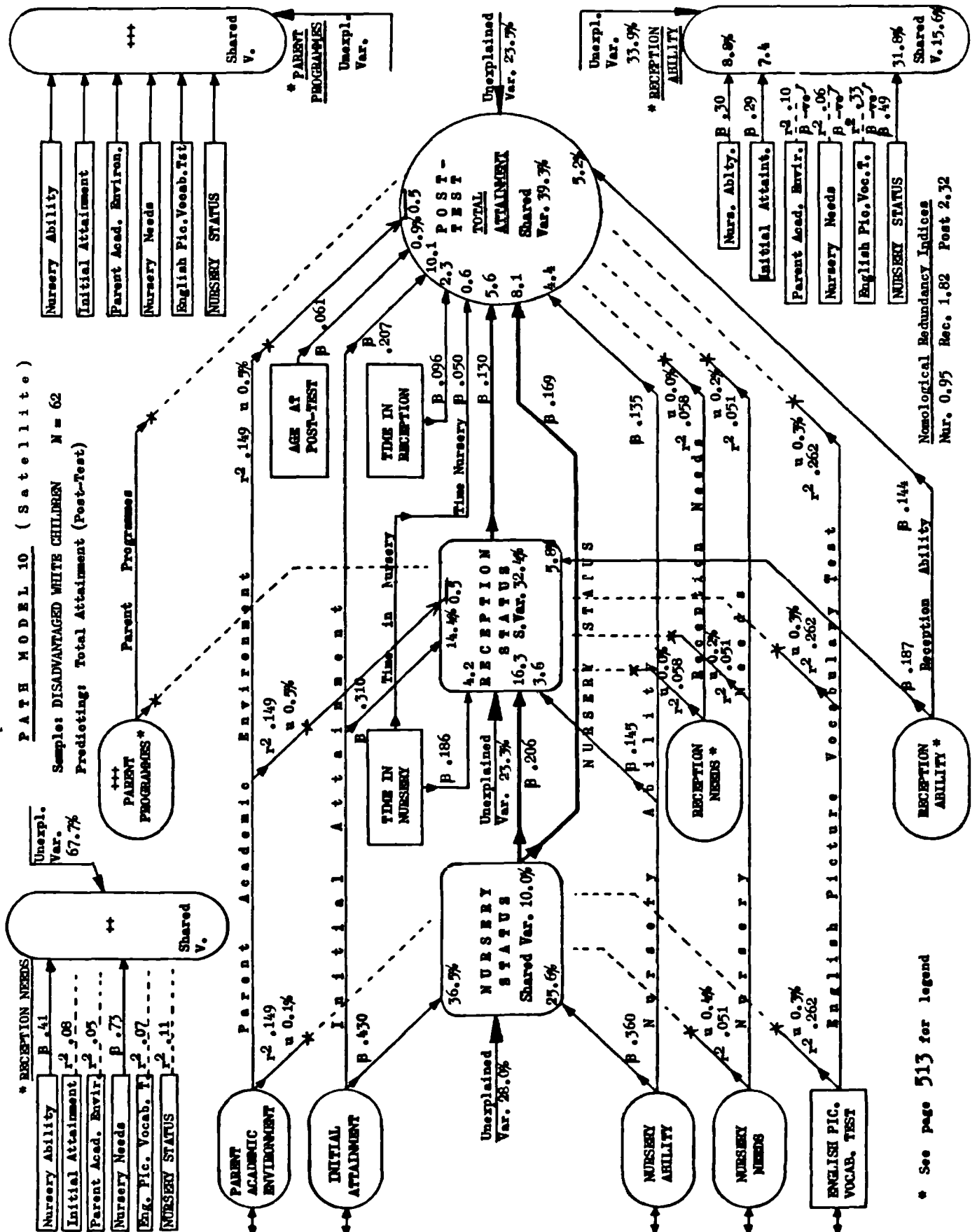
The regressions on the various Needs variables show that for White children these variables predict between two and three per cent of outcome variance; for Blacks the predictive power of the Nursery Needs variables on their own is 15 per cent, and for Reception Needs variables on their own is over 19 per cent. Again it is useful to surmise why the differences should be so considerable. Is it that the Black child in the Nursery class is more distressed and lacking in esteem in a situation where for the first time in his or her life it is Whites who are



in control of the daily environment, with a different dialectical language structure and different cultural perspectives and priorities? Or is it that the Nursery staff have interpreted ethnic and cultural differences in terms of Need for Security and Need for Esteem? The former rather than the latter argument may be the more credible explanation.



<u>Nomological Redundancy Indices</u>		
Mar. 1.29	Dec. 1.93	Post 2.50



\* See page 513 for legend

### Models 9 and 10: Construction of Path Models

The following points can be noted, in comparing the regression parameters for the Disadvantaged Black Children's sample with those for the Disadvantaged White Children.

1. The total outcome variance explained in the Black model rises from 52 per cent at the Nursery level to 63 per cent at the Post-Test level, whereas the variance explained in the White model rises from 72 to 76 per cent. There are various possible reasons for this considerable difference in predictive patterns. The measures used at the Nursery level may not have been sensitive enough to the characteristics of Black children - although in fact the differences in mean levels of performance between White and Black were minimal. Alternatively the Blacks may have included more 'later starters', children whose skills only become apparent as they passed through Nursery and Reception classes.

A third explanation, which incorporates the previous explanation, is that the Nursery and Reception experiences are of particular academic value for Black children, introducing them to the cultural expectations of the mainstream community and this in turn giving them the confidence to learn the skills taught within these two school settings.

2. The nomological redundancy indices for the two models are reasonably satisfactory, with the Post-Test level indices rising somewhat above 2.

3. The pattern of differences between the contributions of Initial Attainment and the Ability latent variables reinforce the differing interpretations offered in the discussion of the first item above. For Black children Ability counts for more than Initial Attainment, though the gap is not wide. For White children on the other hand Initial Attainment accounts for a good deal more outcome variance than do either of the Ability variables.

4. The differences between the predictive power of Parent Academic Environment across the two models, when the constituent variables are examined in isolation, have already been described. Surprisingly the variable makes hardly any contribution to outcome variance, for either the Black or White groups. However there is a noticeable difference in its squared correlations with the outcome variable; for the Black group it is 0.025 and for the White group 0.149. (For the latter there is a very small predictive variance of 0.5 per cent at the Reception and Post-Test levels, but not at the Nursery level.)

Putting together these various pieces of evidence on the size of the variance predicted, on the differential patterns of Ability and Attainment, and on the differences in the parent environment variable in the two models, it is possible to suggest that Black children may start school with a degree of academic handicap - facing a new and strange ethnic environment, and possibly not having

been prepared for the demands of the mainstream community in the same way as are many of their White peers. This is not a matter for blame - either of the school staff or of the Black parents. It was E's experience that on balance Black children tended to give more care to their younger siblings than did White children; there were other warm and encouraging features of Black family life and a deep parental concern about the school prospects for their children.

The fact that the model shows it is Ability which counts more strongly than Attainment for Black children does point to a greater challenge for them, having to rely more in inborn ability and possibly having had less of the needed home academic experiences than many White children. There was no evidence in the study of noticeably poorer performance by Black children, and in fact the performance of some of the brightest Black children - in particular one Black girl who had the highest Maths scores of any of the children in the five Disadvantaged schools - shows that they do not suffer from the inherent intellectual limitations which have been ascribed to them by critics such as Jensen (1969).

But to say this is not to ignore the fact that the Black 'home' academic preparation is not always equal to that given in many comparable White homes. Whether it is the task of the home to provide this kind of preparation may well be a debatable point; if however many parents do insist on providing an academically oriented foundation - with parents at the advantaged school giving a very considerable amount of 'academic' preparation before their children reach school - then any section of the community which fails to provide this kind of foundation will inevitably suffer in the highly competitive environment of education, particularly in the later school years. Given the prejudice which Black children have to face in any case in the wider society, it is not unreasonable to argue that a strong academic foundation in the home could only be an asset for their later schooling and careers.

This viewpoint is not necessarily that of other students of Black educational performance.

On the one hand Sowell (1976, 1980), a leading Black educationist in the United States, offers some profound historical insights into the education and other achievements of Blacks in that continent over the past few centuries. What is particularly important in his analysis is that he shows that most of the great Black leaders of the past and present came from a relatively small number of identifiable schools and colleges which were largely Black in their intake and headed by forceful Blacks who recognised the needs of the Black community more clearly than most White heads of integrated schools. Sowell is not hostile to the concept of racially integrated schooling, but he emphasises its limitations in creating situations where the Blacks become the minority group and their interests take second place to those of White or other groups.

Sowell questions the belief that it is segregated schooling which is responsible for poor Black performance and lower I.Q. levels. He considers that it has been common for minority ethnic groups in the United States to show I.Q. levels 15 points below that of the dominant ethnic group at some stage. Even the higher educational attainment of Black females, often cited as being linked to the Black 'matriarchal' society dating back to the days of slavery, has its parallels in other low I.Q. groups. Other myths about Black educational performance can also be questioned by showing that similar problems have arisen in different U.S. minority groups. What is needed for Black education is not more innovation and huge increases in funding, but a clearer perspective on what are the key educational issues for Black children and how these should be tackled.

While the parallels with the situation of the Blacks in the United Kingdom are not too close, largely because the West Indian and Asian communities have settled relatively recently in this country, the issue of Black educational progress is one of considerable concern, politically and administratively.

A rather different view of Black education from that of Sowell is offered by Taylor (1981) in her study of the factors underlying the achievement of West Indian children in the United Kingdom. She considers that a large body of evidence does point to their relatively poor performance both in academic attainment and cognitive skills, when compared with White or Asian children of similar socio-economic levels. Her study was conducted on behalf of the Government's Committee of Inquiry into the Education of Children from Ethnic Minority Groups.

Taylor recognises the multi-dimensionality of the problem and reviews the considerable amount of research into this matter in the United States, Britain and the West Indies. Her discussion is, however, focused on two fields of inquiry: the educational attitudes and the culture of the West Indian parent and child, and the attitudes, culture, policies and practices of the schools, both here and in the West Indies. She finds it puzzling that West Indian parents are by and large highly ambitious about their children's education and future careers, while the children's achievements do not appear to match these expectations. In comparison there is a closer match between White and Asian parents' ambitions and the achievements of their children.

In historical terms her study also refers to cultural and attitudinal factors within the West Indian Colonial situation, but pays little attention to the centuries of negation of the parental educational role within the former slave culture.

In the light of all these considerations Taylor moves to examine the school environment and teachers' attitudes towards West Indian children, their achievements and their culture. Broader educational policies are also reviewed.

Her conclusions are that more should be done to change attitudes, policies

and practices within the schools, and more efforts made to help West Indian parents to achieve a greater understanding of the school process and the teacher requirements. "If the school can rebuild the trust and confidence he (the child) needs, both by the example of its caring relations and pattern of behaviour which it establishes.... such a pupil may become both self-confident and disciplined and have an accurate self-image, so that he may come to achieve.... In so doing the school will also have rediscovered its own particular educational viability so that it may at last be possible for the high hopes which many West Indian parents have their children to be fulfilled."

While Taylor refers briefly to the Haringey Reading Project (Tizard et al, 1982), showing how West Indian and other parents could make a major contribution to the improvement in their children's reading levels through a daily session of listening to the children's reading at home, in general her study fails to give any serious attention to parental 'behaviours' and the importance of what the parent does for and with the child, in educational terms, at home. Her almost exclusive focus on attitudes and cultural differences serves to minimise the overwhelming importance of practice and exaggerates the importance of responses given to researchers inquiring into 'attitudes towards education'.

In contrast, the evidence described in the present study suggests that attitudes towards education, as assessed in several attitudinal questions in the parent interview, not only fail to predict any part of the variance of educational performance at the end of the Reception year, but also have no meaningful correlation with performance levels - in other words, attitudes are not even subsumed within or represented by the measured parent behaviours in the fields of reading and mathematics.

A somewhat disturbing aspect of Taylor's rather limited focus on attitudes and cultural origins is that, while there is much research evidence confirming the existence of prejudice within the host society and in many of its institutions, it is the schools which tend to get the principal blame for the lowered performance of the West Indian child. While Taylor expresses the aim of defining the positive conditions of home and school life which promote achievement in West Indian children, her conclusions point clearly at the role of the school rather than towards those of the school and home. The massive differences in achievement between pupils of West Indian and Asian origin are ascribed mainly to the fact that the Asian culture is relatively isolated, proud and independent, whereas the West Indian culture is said to be based in part on an outdated model of the host culture, with the host showing insufficient respect for the West Indian, her culture and her rights. There is sound evidence supporting such assertions, much of it cited in Taylor's study. But the equally important part played by what the parent does or fails to do in the home is virtually ignored.

To sum up, Taylor's study adds support to the view that education is only about schools and teachers, with parents seen as people who need to be supportive and aware of the important professional role of the teacher. The modern view that education is a matter of joint effort between teacher and parent, with the parent seen as the principal pre-school teacher and the continuing home 'ancillary' during the schools years, is not reflected in her review.

Even more serious, if the cultural and attitudinal adaptations were to be made as suggested by Taylor, with continuing modifications of school policies and practices followed by yet further findings of poor West Indian performance, society and its research focus will be turned back to a renewed examination of the discredited hypothesis of intellectual inferiority (Jensen, 1969, for example). That would be a retrograde development, since the more credible explanation, that poor performance is rooted as much in parental behaviour patterns as in the schools, has yet to be researched in detail. The Taylor hypothesis, given its importance in the context of a Government Committee of Inquiry, offers only an institutional approach to a grave educational problem. A definitive review of the West Indian child and her real educational potential has yet to be written; the parents and their behavioural environment, in its school and academic orientation, will be an important part of that review.

5. E.P.V.T. continues to show moderate squared correlations with the outcome variable (0.17 for Black children and 0.26 for White children), but makes no contribution to variance predicted. The reasons for this have been discussed in earlier models.
6. As already explained, Parent Programmes have not been retained as a variable in these models.
7. Model 9 is the only model in which Nursery Needs plays any meaningful part as a contributor to outcome variance. While the totals are small - 2.2 per cent at the Nursery level and 0.8 and 0.7 per cent at subsequent levels - it does suggest that this measure provides explanatory information which is not yielded by any other variable. The implications of this finding is that the Black children's Nursery Needs - consisting mainly of the Nursery staff's assessment of the children's lack of self-esteem - does offer a useful though small

addition to...../



addition to understanding of the variables predicting post-test performance by these children. Surprisingly, Reception Needs makes no contribution to either model.

8. Both Time in Nursery and Time in Reception make meaningful contributions to outcome variance in each of the models. The contribution of Time in Nursery is slightly higher for Black than for White children. This confirms what has been said earlier in the present discussion, as well as in the discussion of earlier models, about the fairly strong evidence of the value of the early school experiences for the disadvantaged children in this sample.

9. Age plays a useful part both in the creation of the Ability latent variables and also as a contributor at post-test level. For Black children Age at Post-Test predicts 2.5 per cent unique variance, compared with only 0.9 per cent for White children, while the contribution of the earlier age variables is also a little higher for Blacks.

#### Subsidiary regressions

The regression of Reception Ability on the Nursery level latent variables does not yield any differences of interest.

However the regression of Reception Needs on the Nursery level predictors offers interesting additional confirmation of the hypothesised relationship between Needs and Attainment in the context of the Black child's school experience. Whereas the Nursery predictors for the White children show Nursery Needs and Nursery Ability as the only meaningful predictors of Reception Needs, for the Black children the meaningful predictors are Initial Attainment, Nursery Status, Nursery Needs and Nursery Ability. The contribution of Initial Attainment in this regression is the highest in any model, being over 12 per cent out of a total predicted variance of 40 per cent. What this implies is that the Black children's attainment at Nursery level is a powerful predictor of their later level of Reception Needs (Needs being scored negatively, as pointed out earlier).

#### Overall findings

The following are the key findings which can be drawn from the Disadvantaged Black Children and Disadvantaged White Children's models, subject to the usual caveats in regard to the sample, the data and the analytical algorithms used here, and the employment of satellite models rather than independent models.

a. A much higher level of variance in the outcome of Total Attainment can be predicted in the White children's model than in the Black children's model; the level of prediction in the Black model is relatively low at the Nursery stage of assessment.

b. While Ability counts for more of the prediction of outcome variance than does Attainment in the Black children's model, the reverse is true for the White model, where Attainment accounts for a good deal more of the explained variance.

c. The Parent Academic Environment does not predict any meaningful amount in either of the models; this is in accord with the findings in the Total Disadvantaged Sample model. However the squared correlations with the outcome variable and in particular the regressions of the outcome variable on individual parent academic behaviours show that whereas White parents' behaviours have some identifiable relationship with the Total Attainment outcome, the parallel relationship with the Black parents' academically oriented behaviours are much lower. The possible reasons for this have been discussed at some length in the preceding text. Although it may be argued that since this variable serves no predictive function in either model it is not really important, on the other hand the very strong prediction of Parent Academic Environment in the Advantaged Sample (Model 2) shows how potentially important it is. Even within a disadvantaged sample one may reach a tentative conclusion that many of the characteristics of Black homes, though strongly social and communal - such as the degree of caring by older children for their younger siblings, and the parental commitment to their children's educational progress - are not as yet sufficiently geared to the academic demands of the school. When the effects of prejudice from the wider society are added to this factor, it can be argued that Black children do start school with some degree of educational handicap. However there is no evidence of intellectual inferiority and in fact the evidence from the path models show that Black children rely more on intellectual skills to achieve at school than do their White peers. The early academic progress of the White children is more dependent on their initial attainment levels.

d. For Black children the Needs variables appear to be more predictive than for White children. There are a number of statistical factors which support this conclusion, apart from the fact that Black children entering a White dominated environment for the first time in their lives are more likely to feel concern and a lack of self-esteem in what is for them a doubly strange environment.

e. Time in Nursery and Time in Reception each contribute to outcome variance, in both the Black and White models. The contribution of Time in Nursery to Black academic progress is slightly greater than it is for White progress.

## 6.90 Cost-benefit indicators

The restrictions of time and the relatively modest levels of measurement possible in a study such as the present one have meant that the extensive data needed for a full cost-benefit analysis, as conventionally understood, with discounting of expected future benefits or costs, could not be gathered for a comparative assessment of the cost-effectiveness of the parent programmes. It has however been possible to carry out a relatively simple exercise in which the results and costs of these brief programmes are compared with two alternative programmes, one being a proposed new parent programme, to be run by nursery class teachers or infant school teachers for all the parents of nursery children within schools in disadvantaged areas, and the other being the customary remedial teaching programmes given to children who are found to be poor readers soon after they reach junior school.

Several caveats should be noted. The parent programmes as undertaken for the study are not necessarily representative of the kinds of programmes which might be offered by other workers in this field. Although there are grounds for believing that the parent and child samples were reasonably representative of the parent and child populations in the metropolitan authority area, different results might be obtained on parent groups with a different cultural and ethnic make-up, or with other levels of poverty or relative affluence. Finally, the estimated costs and results of remedial intervention are not necessarily representative of what might be found across a random sample of schools, teachers or children.

## 6.91 Comparison of costs

### Study Programmes

Quantifiable results from the present study, in the five disadvantaged schools, show that for the 48 parents in the reading programme groups and the 32 parents in the mathematics programme groups (with a maximum of six parents in any one group), a total of 190 hours of worker time were spent in administering the 8 fortnightly meetings which were held for most of the groups. The meetings generally lasted one hour. The preparation of programme materials, when costed in time and resource equivalent time for these groups, took a further 120 hours. The initial recruitment of the groups, together with publicity and notification of meetings (but excluding the time spent interviewing the parents), occupied about 20 hours. This yields a total of 330 hours, which can be divided in pro-

portion to the numbers taking part in the reading and maths groups. The time spent on the parent group meetings at the advantaged school is of course not included here.

The estimated 330 hours of time needed for the programme preparation and administration at the disadvantaged schools points to an average of about 4 hours per parent. The contribution in time of the parents themselves is far greater. Individual parents spent a total of between one and eight hours at meetings, as well as an estimated half that time on additional travel to and from the meetings; the time parents spent on programme-related activities with their children ranged from an estimated few hours to 30 hours or more. At a conservative estimate the average parent would have spent altogether 20 hours on attending and carrying out the programme during the eight fortnights that it was under way.

The question arises of whether the time spent on the programme by the parents meant the foregoing of economically viable alternatives. The most important of these possible costs would occur in relation to the likelihood or otherwise of part-time employment as a viable alternative to the one or two hours a week spent on the programme; another quantifiable alternative would be the time that might have been spent searching for bargains at the supermarket, for example, or on other economically-focused activities. Apart from such clearly definable alternatives, it may well be argued that the periods of reading or maths-related interaction between the parent and child were reasonably enjoyable and not necessarily a 'loss' in terms of alternative activities such as housework or leisure time. It is likely therefore that the main costs of the programme were the time spent on organising and holding the meetings, together with the equivalent time spent on resources - a total already estimated at four hours per parent - plus a relatively small amount of parent time in the few situations where programme attendance or child-oriented programme activity might have reduced part-time employment or some other economically beneficial activity. These latter possibilities are unlikely to amount to more than an average of two to four hours per parent.

#### Regular Nursery Programme

The possibility of a formal parent programme, carried out within every school where there are nursery children, appears to be a valid option in the light of the findings of the present study. Such a programme could stretch over the entire nursery 'year', with a total of seven monthly meetings for the parents. Instead of the one-hour meetings which formed the bulk of the separate reading and maths study programmes, the regular nursery programme could consist of two-hour meetings occupying most of a nursery morning or afternoon, having sessions devoted to discussing both reading and mathematical activities or 'games' for the children, and having a maximum of about six parents in any one group

(a typical nursery might therefore be running a fair number of these groups in any one year).

It is not unreasonable to think that such a programme would achieve results as good as or better than the study programme, particularly if there was considerable staff encouragement of parents to continue attending what would then be seen as a 'normal' part of the nursery curriculum. A particular saving in such a programme could arise from the centralisation of the production and distribution of resources by the relevant education authority.

### Remedial Reading Programme

The amount of time given to remedial teaching varies widely from school to school. In some cases children are taken out every day for intensive tuition, in small groups or even individually. In other cases remediation is offered within the classroom setting. Here the example will be taken of a remedial reading programme at one of the schools used in the present study. At the junior level of this combined school, children judged to be in need of remedial instruction are taken into a small group situation - up to six children in a group - twice a week, for an hour at a time. This extra teaching, which is given while the main class is busy with other work, may continue for anything between a year and three years, depending on the difficulty of overcoming individual problems which may have many facets beyond that of reading difficulty itself. An average period of remedial attention could last two years - amounting to 160 hours of special tuition. The division of the hours by the number of children in an average group would suggest that an extra  $160/6$ , or 23 hours of individual remedial instruction is given to each child.

Unfortunately there are no comparable examples of remedial maths programmes at infant or junior school level, suggesting that mathematics is not yet widely recognised as one of the cornerstones of education for participation in a modern society.

Limiting this comparison then to a remedial reading programme, it is clear that it would be difficult to compare the estimated results of any such programme directly with the results of the study's parent programmes. For the latter the multivariate analyses described in section 6.84 offer strong evidence of the effectiveness or otherwise of particular parent programmes and on the predictive strengths of the programmes in the presence of a variety of other predictors. This is a most stringent form of judgement.

For comparing the results of remedial teaching directly with those of the parent programmes it would be necessary to assess a large number of competing predictors in the remedial situation, with a view to presenting a comparable multivariate model in which the unique contribution of the remedial teaching could be

teased out, separately from the contribution of other predictors. Thus, for example, if a remedial reading programme is carried out on a continuous basis at a junior school, with children in the age range 8 to 10 spending between one and three years in the programme, the results achieved will undoubtedly reflect not only the teachers' skills but also the level of home support for what the teachers are doing, the child's cognitive levels, prior attainment, and even emotional state as assessed by estimates of the child's need for esteem and security.

Unfortunately there is usually only a limited amount of data available from remedial programmes, such as the length of time the children spend in the programmes and their initial and final reading scores. These data can, however, be linked to other data from the teaching of reading. For example, at the age of nine years (the median age for a junior child), the child will have experienced approximately four years of instruction in reading. While there is no clear demarcation of class time in most infant or junior schools, it can be estimated that the average child will have experienced about one hour of concentrated reading instruction a day over those four years, giving a total of 800 hours of instruction. The problem for the child is that the daily hour (or series of shorter periods amounting to an hour) has had to be shared with approximately 24 other children. Again these are figures which can vary widely between schools.

The implication of the figures suggested above is that each child will have had about  $800/25$ , or 32 hours of individual teacher instruction in reading by the time it reaches nine years of age.

These three types of programme will now be costed.

To facilitate comparisons the costs are assessed in the form of equivalent teacher hours (E.T.H.) and the numbers of parents in the nursery programme are equated with the number of children receiving remedial help. The assumption of both the nursery programme and the remedial teaching programme is that while not all parents/children would need nor indeed would benefit from the interventions, the programmes and remedial teaching can be of help across the range of competencies likely to be encountered at schools in disadvantaged areas.

Table 40 (overleaf) summarises the costs associated with the different programmes.

Table 40 ...../

Table 40. Cost in equivalent teacher hours of three intervention programmes.

	Study Programmes (3 to 6 per group)		Regular Nursery Prog. (3 to 6 per group)		Remedial Tch. (6 per group)
	48 Reading pr. parents	32 Maths pr. parents	80 parents given comb. reading/maths prog.		80 children, rem. reading
	8 fortnight. meetings (1 hr. each)	8 fortnight. meetings (1 hr each)	7 monthly meetings, 2 hours each		6 school terms, 2 hrs per week
Duration of programme			Reading                  Maths		
Programme meetings (equiv. teacher hrs)	114	76	112	224 112	2240
Preparation prog. materials (E.T.H.)	72	48	-	-	-
Centralised authrty. prepartn. materials	-	-	10	10	-
Recruitment and publicity (E.T.H.)	12	8	10	10	-
Totals (E.T.H.)	198	132	132	132	2240

#### 6.92 Comparison of benefits

The assessment of programme benefits is considerably more difficult than the calculation of costs. As already pointed out, the data available on remedial teaching programmes do not (usually) permit a wide-ranging multivariate and longitudinal analysis. A more basic comparison will therefore be made, using direct prediction estimates based on simple regression equations in the absence of competing predictors. These data are already available for the study's parent programmes, and by definition the same data can be applied to the projected regular nursery programmes for parents. For the remedial programme the statistics presented earlier, together with a generous estimate of potential benefits, will be combined.

#### Study Programmes

In the creation of latent variables for path models 4 and 6, in section 6.84, it is shown that parents' reading programme attendance (at the disadvantaged schools) has a regression coefficient of 0.206 (p 0.013) in predicting the child's post-test reading scores. In effect, an increase of one standard deviation in parent programme attendance could lead to an increase of 0.20 standard deviations in reading scores. For the parents' maths programme attendance the coefficient of prediction of the maths concepts outcome is 0.291 (p 0.006). Thus here an increase of one standard deviation in parent programme attendance

could lead to an increase of 0.29 standard deviations in maths concepts scores. These and other estimates (below) of the effect of a one standard deviation increase in the length of a programme assume a continuity of effect with an extension of the programme. Clearly this would not necessarily be the case if there was a considerable extension, since a longer programme may well change in character or cause more serious dropout problems. However the use of a uniform hypothetical extension of 1 standard deviation enables a tangible comparison to be made between programmes. The comparison would obviously apply *pari passu* to an increase of one-tenth of a standard deviation in the length of each programme.

#### Regular Nursery Programme

It is reasonable to hypothesise that the programme outlined earlier, with seven monthly meetings of two hours each and dealing with both early reading and early mathematical game-like activities for parents to undertake with their children, could be as efficient or more efficient than the study programme. The motivational and other general features which the study's reading and maths programmes had in common would now be combined within the one programme and it could be expected therefore that the 14 hours needed for such a programme would have the same (or a better) effect than the combination of the two separate 8 hour programmes which were organised for the study. It is thus assumed that the predictive coefficients derived from the study programmes can be accepted as the basis for predicting the achievements of the regular nursery programme. As the latter programme is offering both reading and maths guidance to the parents, it is assumed further that the programme children will experience improvements in each subject, equivalent to the improvements achieved in the separate reading and maths programmes.

#### Remedial Reading Programme

It is necessary to translate the figures quoted earlier on the length of normal teaching experienced by children, and the length of time spent on remedial teaching, into equivalent standard deviation measures. The variance of a rectangular distribution (such as the length of a teaching programme ranging from zero to the estimated 32 hours of individual reading instruction which a pupil of 9 years old will have enjoyed) is measured by

$$\frac{(a - b)^2}{12}$$

where      a      is the upper limit of the distribution  
            and      b      is the lower limit.

For the distribution of the normal teaching of reading, this gives a variance



estimate of 85.33, equivalent to a standard deviation of 9.24. The comparison with the estimated 23 hours of remedial teaching (as costed in section 6.91) is surprising. In effect the average remedial child receives just on 2.5 standard deviations of additional reading instruction.

It is difficult to hazard an estimate of how much success or otherwise is achieved with the average remedial child. It is intimately related to a number of factors, inter alia the question of how narrow or broad is the definition of backwardness when deciding on allocation to a remedial group. A generous estimate will be made here. For the purposes of comparison it is assumed that, on a reading test with a mean of 100 and a standard deviation of 15, the effect of a successful remedial programme would be to achieve a mean improvement of 15 points, or one standard deviation, in the reading scores of the remedial children. By any judgement this would be a considerable gain. In effect, this estimate assumes the achievement of an average one standard deviation increase in the reading scores of remedial children in relation to their peers within the non-remedial group in the same classes. Clearly, then, this increase is above and beyond that occurring in the course of the normal teaching process.

The table below presents the comparison of costs in relation to short-term benefits across the three programmes.

Table 41. Hypothesised effectiveness of three intervention programmes.

	Study Programmes		Regular Nursery Prog.		Remedial Tch.
	Reading Prog.	Maths Prog.	Attribution of prog. to Reading	Maths	Reading
Number of people	48 parents	32 parents	80 parents	80 parents	80 children
Hours of equivalent teacher time (E.T.H.)	198	132	132	132	2240
Effect of one stand. deviation increase in programme length	0.20 s.d. improvement in reading	0.29 s.d. improvement in maths	0.20 s.d. improvement in reading	0.29 s.d. improvement in maths	1/2.5, or 0.40 s.d. read. improv.
Est. cost of one s.d. incr. in prog. length (assume rectangular distrib. of extension) in E.T.H.	$198/(12)^{\frac{1}{2}} = 57.2$	$132/(12)^{\frac{1}{2}} = 38.1$	$132/(12)^{\frac{1}{2}} = 38.1$	$132/(12)^{\frac{1}{2}} = 38.1$	$2240/(12)^{\frac{1}{2}} = 258.7$
Est. cost per parent-child dyad (E.T.H.)	1.19	1.19	0.48	0.48	3.23
Est. cost per 1 s.d. improvement in reading or maths per child (E.T.H.)	5.95	4.10	2.40	1.67	8.08

### 6.93 Cost-effectiveness considerations

Even with the very generous estimate of remedial programme effectiveness, the above comparisons suggest that parent programmes are more economic or cost-effective. The relatively high costs of the study programmes, compared to the proposed regular nursery programme, are related to several factors - the considerable time that was needed for the construction of the programme materials, the division of the programme into separate reading and maths groups (as a form of research control), and other organisational problems concerned with the pioneering nature of the work. An education authority programme (the proposed 'regular nursery programme'), with each nursery class running its own programme but having the materials designed and reproduced centrally, and combining reading and maths in one programme, could cost a good deal less, as shown by the above figures.

There are of course ~~some~~ other costs which could be included. For the parent programmes the economic cost of time foregone by the parents - that part of the parents' programme time which could have had alternative economically profitable uses - has not been included in these comparisons because of the uncertainty as to how the time spent might have been assessed by the parents. A far more serious cost has been omitted from the remedial programme, namely the early loss in general academic attainment suffered by the reading 'failures'. For example, inability to read well may have limited the child's advancement in mathematics. A further loss in other learning time would occur when the child was given the additional remedial instruction, whether or not the remediation took place in the classroom or in groups outside.

Beyond this stage of comparison there would need to be formal cost-benefit assessments of children's long-term achievements. The preceding pages have provided rough estimates of the short-term costs and benefits of the programmes. For a fuller picture the benefits would need to be projected forwards in terms of estimating later school functioning, particularly in relation to achievements as measured by school certificates, and in terms of occupational qualifications, when compared with similar children who were not part of any parent or remedial programme. Longitudinal data sets already in existence in Britain and the U.S.A. could be of use in carrying out such projections, which in turn would need to be discounted to arrive at present-day benefits prior to applying the projections to existing parent or remedial programmes.

There are three further considerations. Firstly, remedial success below the estimated high level of a one standard deviation increase in reading scores would increase the gap between the estimated costs of the remedial reading programme and the parent programmes even more. Taking other factors into account,

including those mentioned in the previous paragraphs, the comparisons set out here could well show the minimum gap between parent and remedial programmes.

Secondly, it should be recognised that the samples of children whose parents take part in a nursery programme would include some but not all the children who might later be found in a remedial programme; while the two samples are therefore not strictly comparable, even if the age gap is ignored, there is no reason to assume that a programme of non-selective 'remedial' reading instruction would necessarily be more cost-effective than a programme focused on those at the bottom of the reading distribution. For the latter group the effect of regression to the mean is likely to boost the apparent success of remediation.

Thirdly, in any long-term comparison of parent programmes with remedial programmes, a particular strength of the parent programmes is likely to be the fact, reported in various American studies, that parents frequently continue with their improved child stimulation behaviours once the programme has ended, even if not at the same high level as originally. In contrast, remedial teaching programmes may help to lift the level of performance directly but they seldom have long-term effects on the child in the way that parent programmes are reported to have.

While no final conclusions can be reached on the cost-effectiveness of the present parent programmes, there are grounds for claiming that a school-organised parent programme, making use of centrally produced resources, could be more efficient in organisation and resource costs than the programmes described in this study, and considerably more effective than any later remedial teaching programme. To spell out a practical contrast one can point to the difference in teacher requirements for parent and remedial programmes. For an infant school with two nursery classes (say, morning and afternoon sessions), having a total of 50 mothers (or fathers) willing to attend the monthly parent meeting - even where it means taking off a few hours of working time - a total of seven weeks of teacher time would be needed for organising and running the monthly classes for an estimated 10 groups over seven months. To achieve the same results, in terms of an equivalent improvement in standardised reading and maths scores for the same number of children, would require about 27 weeks of teacher time.

The earlier work of Hewison and Tizard (1979), the present study and a growing number of American studies all point in the same direction, namely that the parent teaching resource is the most economic resource available to the school system in the early years of a child's life.

## 7.00      S y n o p s i s

This final chapter will first note some major conclusions drawn from the study's review of the literature on early educational attainment and intervention. This is followed by an outline of the study itself, its exploration of some new methodological and statistical approaches, the strengths and limitations of its design and analyses, and its principal findings, especially in relation to the formal hypotheses. Some tentative conclusions and possibilities for the future are put forward.

### 7.10      Summary of previous chapters

The paragraphs below present brief abstracts of the most relevant parts of previous chapters, with added comments at certain points. The bracketed numbers following each abstract refer to the sections from which the summary is taken. Authors' references are not provided here, but appear in full in the numbered sections.

#### From Chapter 2:      Early Education Attainment: a conceptual model

Some reading theories.      One of the most important theorists on the development of reading, Frank Smith, has shown that the evidence on reading acquisition 'suggests that the only way to learn reading is by reading (just as the only way in which a child learns spoken language is by talking and listening)'. In contrast, research on the Frostig hypothesis has shown overwhelmingly that training in perceptuo-motor skills does not lead to a concomitant improvement in reading attainment. These and similar findings may call into question the value of the practice of giving nursery children perceptuo-motor activities in preparation for infant school, rather than encouraging the children to practice recognition and drawing activities related to letters, words and numbers. (From section 2.11)

The case for early reading.      In general schools offer little or no guidance to parents on pre-school reading activities, suggesting instead that parents should read to the children and offer more 'language experience'. There has been an educational emphasis on the 'harm' which can be done to young children by encouraging them to read too early, despite a lack of evidence for such claims.

Southgate's examination of the reading debate is accompanied by a detailed rebuttal of every argument put forward against early reading; studies have shown that slow learners and even brain-injured children can be taught to read at ages from three upwards; to spend a short time each day enjoying the game of reading does not detract from other activities. Durkin too has shown that the lower the intelligence level the greater the advantage of an early start with reading. (From sections 2.13-2.15)

The development of early mathematics. There is a much stronger sequentiality in the acquisition of mathematical attainment than there is in reading, where a number of pathways can lead to literacy. There is also possibly a greater maturational element in mathematical attainment, though the extent of this is disputed. Elkonin considers that the psychological processes underlying the learning of mathematics are only formed in the process of learning, rather than preceding it; the cognitive development of children can indeed be altered by school instruction in mathematics. (From section 2.20)

Influences on ability. Among the factors identified as contributing to a child's ability, and indirectly therefore to its later academic attainment, are: nutrition, both in pregnancy and in the first years of life (with malnutrition, in contrast, leading to apathy or even brain damage, so that the child may not respond adequately to environmental stimulation); culture, which is closely interwoven with cognition, so that major differences in the cultural environment may lead to qualitative differences in thinking; play, which can be of value in fostering the perceptuo-motor and social skills of the child; and motivation, which is exemplified by the temperamental or other characteristic approach of the child to exploration and problem-solving but which is also conditioned by its earlier successes and failures. The nature of the interaction between the environment and the child's cognitive-motivational functioning is of much greater importance than cognitive functioning on its own. (Section 2.31)

Damaged and delayed development. Hunt suggests a mutual interrelationship between poverty and lack of competence - competence seen as the abilities, motivation and conduct required for reasonable academic success. Research has learned that poverty contributes indirectly to the impoverishment of verbal communication at home, insufficient or incorrectly sequenced reinforcement during interactions in infancy, and a lack of regular temporal and spatial organisation in daily living. (Section 2.32)

The influence of temperament. Studies have shown that the temperamental characteristics identified soon after birth are of considerable importance in determining later behavioural responses to the environment and to experience. The genetic environment as well as the neurological and nutritional influences

experienced by the child prior to and during birth contributes to these behavioural patterns, and so does later interaction with the environment. A variety of specific temperamental or personality characteristics have been identified as contributing to attainment. These include attention span, cognitive style (impulsivity/reflectivity and field dependence), self-concept and locus of control. It is also possible that attainment and abilities themselves help in some measure to influence the development of personality. (Sections 2.41, 2.42)

Motivation. Bruner considers that intrinsic motivation - the will to learn - is of key importance in learning. Intrinsic factors include curiosity, self-directed attention, interest and learning structure. There is however a considerable difference between the approaches of those who consider that the child needs to choose methods and materials reinforcing to her or himself, and the view of others that, for disadvantaged children at least, it is structured educational experiences which enable motivation to develop. This difference is important in relation to the kind of curriculum offered in the pre-school and early school years. Some research has also suggested that consistent, firm and loving discipline in the home is most likely to produce independent, self-controlled, exploratory and competent behaviour in the child. (Section 2.43)

The social milieu (maternal employment). The claimed effects on the child of the mother's employment is a much debated issue. Some evidence shows that mothers employed outside the home are more likely than other mothers to encourage independence in their children and to show other positive child-rearing behaviours. In the absence of adequate statistical analyses it is uncertain whether or not these findings are merely a reflection of the occupational levels and 'upward mobility' of those women within any one class who do succeed in finding work outside the home. While some writers urged that it is preferable for mothers of pre-school children to be at home caring for them in the early years, other writers argue that in the interests of both mothers and children, institutional care should be provided for all children from 3 years onwards. There is considerable pressure by mothers themselves for more nursery and other day care facilities for their pre-school children. (Section 2.51)

The social milieu (family income). Margaret Wynn's 1972 analysis of the economic plight of parents points to the persistent social and fiscal neglect of family interests in the United Kingdom over the past decades, with resources being diverted away from parents towards non-parents. Economic forces generally tend to develop adversely to the interests of families if not checked. Wynn urges the payment of a wage to mothers, giving the mother greater freedom to decide on how the child should be cared for, and also giving the State a greater right to ensure minimal levels of emotional and developmental care for infants. (Section 2.51)

Parent-child interaction. Among the many different theories about the nature of parent-child interaction, it is now generally agreed that the child's role in initiating and terminating interaction is as great as or greater than that of the parent. Farran and Haskins consider that the differences in the time spent interacting with the children are a function of the value systems within the social group. 'For a parent to value playing with children as an acceptable adult activity probably requires both the time free from other concerns to engage in such activity and exposure to role models that sanction playing with children as a worthwhile activity.' (Section 2.52)

Child-rearing norms (parental involvement). Urie Bronfenbrenner's classic study of the differences in child-rearing norms between Soviet and Western societies has shown a lack of parental involvement, companionship and intervention in the lives of children in the U.S.A. and Britain, compared with a far more conformist but also more caring Soviet society, with greater cooperation between parents, children and the community. Bronfenbrenner has also concluded, on the basis of research, that any appreciable, enduring improvement in a child's development can be effected only through an appreciable, enduring change in the environment and behaviour of the persons intimately associated with that child on a day to day basis. Even seriously disadvantaged parents can be motivated and provided with the skills needed for developing their children, provided that they are given relief from the burden of sheer survival. He considers that ensuring a high level of expertise in persons dealing with the child (in the institutional setting, for example) may not be as critical for furthering the child's psychological development as creating possibilities for those who are potentially the most important influence - parents and friends - to realise their potential. (Section 2.53)

Child-rearing norms (home visiting). Bronfenbrenner also finds that parent education alone, without involving the children or having them present at the same time, has not been shown to be successful. It is essential, moreover, that in any intervention visits the mother should be asked to play the primary role so that her relationship with the child is not undermined. There should be home visiting during the first three years of life, with pre-school education and group care thereafter. Group care in the first few years has not been found to be effective. (Section 2.53)

Child-educating norms. The Douglas study showed the importance of the parents' educational environment - the mother's aspirations for their children, the parents' visits to the schools, and the encouragement given to do school work; children with interested parents pulled ahead whatever the level of their initial starting ability. On the other hand, research by Tizard has shown that what teachers expect from parents, in relation to 'parent involvement', is not

necessarily the same as what parents expect from teachers. Unfortunately all studies of parent-teacher cooperation have been premised on the view that it is the parents who should enter the teachers' territory and contribute in some way, under the guidance of the teachers, to school life. Parents may however have a more significant educational role if they are given skills which they can apply to their children in the home setting, particularly skills in relation to early reading and mathematical development. (Section 2.54)

The parent as teacher. Hewison has established that after taking account of maternal behaviour and intelligence, the amount of maternal coaching which the children receive is a highly significant contributor to reading performance in the early years. Tizard, Schofield and Hewison in a later intervention study have shown a marked improvement in reading performance for school classes which initiated a parent involvement programme where parents undertook to listen to their children read every evening. Other studies have shown that mothers of children from two years upwards can start teaching them simple reading skills, provided that the child enjoys these programmes as games rather than as a serious learning activity. In general, research and practical initiatives suggest that parents of pre-school and infant school children may have a far more active educational role than is usually ascribed to them. (Section 2.55)

The history and provision of pre-schooling. Among others, Austin shows that although created last century for the benefit of the poor, early childhood education programmes have gradually been adopted by the middle class in six out of eight Western countries, with provision for the poor being diminished and the orientation of the programmes changed from a concern for welfare, health and education, to a concern for the middle class priorities of social and emotional adjustment and creative expression. It is only recently that some intervention programmes have endeavoured to reverse this trend. Nursery policy in Britain has tended to be permissive and unstructured; one of the early pioneers, Robert Owen, laid down the rule that children were not to be 'annoyed with books'. State provision of nursery education in this country has been a story of dramatic ups and downs, with major reversals of policy at intervals. Nursery provision in 1980 ranged from 14 per cent of 3 and 4-year-olds in rural areas to 38 per cent in densely populated urban areas. This total of some 450,000 pre-schoolers is cared for in nursery schools, classes and day nurseries, while a further 600,000 are cared for in playgroups, with an unknown number being looked after by both registered and unregistered child-minders - possibly between 100,000 and 200,000 children. Many children may pass through several different forms of provision during their pre-school years. (Section 2.61)

The role and functions of pre-schooling (U.K. developments). The Plowden Report of 1967 considered that the educational, social, health and welfare



grounds for nursery education were strong. However detailed arguments in the Report focused almost entirely on the socialising and welfare functions of pre-school provision, with only a passing reference to the possible educational benefits of compensating for social deprivation or improving language development. Some 15 years later Lady Plowden declared that she now felt her previous commitment to the expansion of nursery education to have been wrong and that playgroups should be the most important avenue of pre-school provision for the future, as they involved parents closely and were a cheaper form of provision. A separate study of the views of nearly 600 nursery teachers found that most of them gave top priority to the social rather than intellectual objectives, even for disadvantaged children. Woodhead sees that finding as predictable in terms of the history of nursery education and its underlying philosophy - especially as it has distanced itself from the infant school model. (Section 2.62)

The role and functions of pre-schooling (play and structure). There is strong evidence to suggest that structured methods in the nursery classroom have achieved better results, in terms of attaining higher academic skills in the school years, than have the traditional unstructured nursery activities. Barbara Tizard considers that nursery education today provides much less of the kind of informal teaching which was given in Robert Owen's or Margaret Macmillan's nursery centres. The difficulty with the current practice of adapting what is taught to the needs and sub-culture of the child is that it leaves to the child the determination of educational aims, rather than to society, the parents or the teachers. Tizard criticises in particular the many untested assumptions in regard to the value of unstructured play. There is no solid research for most of these assumptions. With the exception of the relatively rich sectors of society, mothers see an almost total mismatch between the values of the nursery school and the working world which their children must one day enter. According to Tizard many nursery schools have failed to provide a situation in which children can both play and be involved in goal-directed activities. Another shortcoming is the failure to facilitate interaction with adults in adult-initiated play, even though watching and copying adults has always been a central educational method. Bronfenbrenner offers many arguments similar to those of Tizard, adding the warning that children who from an early age are cared for in group settings for most of the day are more likely than other children to engage in egocentric, aggressive and anti-social behaviour during the pre-school years and later. (Section 2.62)

Research findings on nursery education (effects). Chazan's review of research shows that nursery education has only modest effects on social and communication skills and early educational attainment. In general there has been very little research in this area of education, with the assumptions and sayings of the early

pioneers and more recent advocates achieving the status of research findings in their influence on thinking. What research there has been suggests only limited and short-term effects from the unstructured or only loosely structured play situations in which there is little adult involvement or guidance. (Section 2.63)

Research findings on nursery education (focus on the disadvantaged). If the current policy in nursery education is essentially that of aiming to facilitate play and happiness, with social, linguistic and educational progress seen as inevitable concomitants rather than deliberate goals, there may be little reason for questioning the reversal of position by Plowden in favour of playgroups for the majority, with nursery investment focused only on disadvantaged children. The nursery school curricula could then be adapted so that the indirect goals of social, linguistic and educational progress could now become direct goals focused on overcoming the developmental and educational limitations of disadvantaged children. (Section 2.63)

The influence of television. Research on the American children's TV programmes 'Sesame Street' and 'The Electric Company', and other research on the English programme 'You and me', indicates that these programmes do have an educational impact and contribute to later attainment. However there is also evidence that the interaction between such programmes and the viewing environment (parents or teachers) may be more important than the programmes themselves. (Section 2.70)

An integrated model. The paucity of integrated models of early educational attainment may be due to the difficulty of assembling all the variables of importance and assessing these competently. Another problem is that studies in this area are often based on a dichotomous model in which the learning disabled or 'backward' children are mistakenly seen as different from 'normal' children, instead of recognising that all children are on continuums across a range of predictive and outcome variables. The analyses carried out on models of attainment also tend to be timid and offer a restricted view of what is a wide and complex picture of relationships. An ideal model is hypothesised here in which different sets of predictor variables are examined sequentially - pre and perinatally, in the infant years, in the pre-school years and in the primary school years. (Section 2.80)

From Chapter 3: Intervening in Disadvantage: home and pre-school

The wider social environment. The close links between social disadvantage and educational backwardness have been identified in many population studies. Among the key factors are the effects of cumulative educational deficit building on

early deficit, a limited attention span, disciplinary problems, unmet physical needs such as hunger, the surrounding community's apathy towards education, and the school ethos in relation to the community. Yet many children succeed despite a background of serious disadvantage, and many families escape each generation from the 'cycle of deprivation'. These issues are linked to questions of how equal is the educational opportunity offered to all children, and whether it is possible to make equal use of that opportunity given the differing early backgrounds of children. The evidence suggests that class factors still play a considerable part in determining educational attainment in Britain, as Halsey and others have shown, whereas meritocracy is a more likely determinant in the U.S.A. (From section 3.12)

The school environment. A large-scale study by a team under Chazan, examining the progress of infant children in three United Kingdom cities, found that children in deprived areas were under-functioning at school in relation to their non-verbal intelligence. There was also evidence of a fair minority reaching school with behavioural problems. Differences were noted across areas in the extent to which parents prepared or failed to prepare their children for school. Within the schools themselves various studies have identified the effects of teacher expectation on the performance of children. A recent study by a team under Rutter identified school ethos as an important factor in predicting both behaviour and attainment. (From section 3.13)

The reality of intervention. Bronfenbrenner points to the stark contrast between a Western world in which the interests of the individual are paramount, with increasing age segregation and alienation of the young, and an East European world in which there is a loss of individual liberty but strong social cohesion and awareness and an intense focus on children and their growth and welfare. Going beyond Bronfenbrenner's thesis, it may be argued that since individualism is a dominant characteristic of Western society, intervention, and in particular pre-school intervention, may have become necessary to minimise the damaging effects of a system in which laissez faire economics play a stronger part in determining progress than do the economics of social engineering. The damage takes the form of poverty, unemployment and communal disintegration for an important minority, contributing to a parental inability to cope with the demands of child-rearing and support in their subsequent education. (Section 3.20)

Head Start: the American dream. While most pre-school education in the U.S.A. prior to 1965 was provided for and geared to the needs of middle class children, Head Start, a vast and ambitious programme set up in that year, aimed to provide a combination of social services, health, nutrition and education for disadvantaged children. The educational component had to rely on rather vague theoretical models, however, as there were few practical models to follow at that

stage. Westinghouse/Ohio produced a highly critical evaluation of the results of Head Start, but ten years later Lazar and Darlington gathered the findings from twelve of the most prominent university-researched Head Start programmes and showed that on every major criterion the relatively large samples of intervention children were in later school life still well ahead of the control children, particularly in academic attainment. The main flaws in Head Start were the allocation of funds in many districts without adequate checks, proper planning or evaluation. On the other hand programmes which were launched with clear goals and formal evaluation generally showed moderate to good results. (Section 3.21)

Educational Priority Areas: the British compromise. Following the Plowden Report, a set of limited budget intervention programmes were piloted in the United Kingdom, focused on schools in areas newly defined as Educational Priority Areas (E.P.A.) - areas based on a set of indicators of social disadvantage. Language and mathematically-oriented school programmes and a home intervention programme were included; results were, on the whole, rather disappointing. The E.P.A. criterion itself has been criticised by Acland and others on the grounds that funding was based on identifying schools in areas of need rather than on the proportion of under-achieving children in all schools. It was also argued that compensation should be based on the need to improve a child's academic achievement, rather than on the diffuse social aims set out by Plowden. (Section 3.22)

Three landmark studies. Three intervention studies of particular importance are those of Heber and Garber, Weikart and colleagues, and Hewison and colleagues. The first named study, known as the Milwaukee Project, succeeded in raising the I.Q.s of a very small sample of 20 severely disadvantaged Black children by an average of 30 points, compared to a control sample, after years of full-time daily intervention with the children. The second study, the Ypsilanti Perry Pre-School Project, followed approximately 60 intervention children throughout their school years; these children, of low socio-economic status and low initial intelligence scores, were given an intensive two-year pre-school experience, while their mothers were visited at home weekly by the children's teachers. Compared to a control sample, the intervention children showed an increasing academic advantage over controls. The third study, that of Hewison and colleagues, has been described briefly in section 2.55 (and earlier in the present summary); the team found a marked improvement in reading scores for young children whose parents were persuaded by the schools to listen to the children's reading every evening at home. While the Milwaukee Project has little practical relevance, because of its extremely high professional input, both the Hewison and Ypsilanti Perry studies are of considerable importance. However the detailed review of

the Ypsilanti evidence, presented in this section, suggests that the importance of the home visits as contributing considerably to the success of that project may have been seriously underestimated by the research team. (Section 3.23)

Institutional intervention (programmes). The focus of most intervention has been on what the institution can do to overcome the believed limitations in the home rearing of disadvantaged children, rather than on stimulating and guiding parents to improve their rearing patterns. Inevitably many of these institutional programmes, when undertaken on a large scale, have involved political and administrative considerations. There is little clear evidence on the effectiveness of such programmes in bringing about long-term improvements for the children. The characteristics of the programmes have been shown to vary considerably, and can be categorised as permissive enrichment, structured cognitive, structured informational and structured environment - the last defining a Montessori-type programme. The structured cognitive and structured informational have been found to be the most successful. Many of the Head Start centres were based on permissive enrichment principles. (Section 3.24)

Institutional intervention (language). Programmes with a clear language orientation offer a considerable challenge since language embraces not only ability and motivation but is also strongly related to cultural factors. Marion Blank and colleagues have developed structured programmes which emphasise the usefulness of the child's previous experience in the questions put and the answers evoked. In contrast Joan Tough rejects the structured language approach, arguing that it neglects the needs of the child to reflect on and use its own inner knowledge to develop a wide range of language strategies. These differences over the importance of structure when intervening with the disadvantaged reflect a fundamental polarity of options among specialists in this field. (Section 3.24)

Institutional intervention (class and structure). Becker and Carnine argue that schools tend to be designed for middle class children whose parents teach them the needed language skills at home. Thus formal language teaching is not a feature of most schools, although in fact disadvantaged children would benefit by conscious language instruction at school, particularly as this could help to overcome their vocabulary deficiencies. These and other authors urge the need for criterion-oriented or 'direct instruction' teaching. Woodhead points out that many nursery teachers in the United Kingdom reject this structured 'American' approach to intervention, although nursery assistants are more sympathetic to it. The contrasting unstructured approach, which holds sway in much nursery and kindergarten practice in the U.S.A. and Britain, providing simply an enriched environment replete with equipment, toys and work materials, is heavily criticised by Williams; in effect, adult intervention has been increasingly depreciated and

finally prohibited in such settings. (Section 3.24)

Parent intervention (professional views). Health visiting in the United Kingdom offers the earliest example of large scale parent educational intervention, having started in the middle of last century with visitors appointed to give advice to disadvantaged mothers on the health and development of their infant children. In general the other professions concerned with children, such as teachers and psychologists, are reluctant to enter the home. Barbara Tizard expresses the reservations of a number of leading psychologists, including herself, about the wisdom or value of home intervention. It is seen as an intrusion, possibly damaging to mother-child relationships, and less effective than group discussions. Changes in the mother's life may be more effective in altering her communication with her children than encouraging her to play with or read to them. Contrasting views are advanced by Tessa Blackstone and some others. Intervention focused on the mother-child relationships may raise the mother's confidence, aspirations and effectiveness. Moreover some families can only be reached by home visitors. One of the problems faced by parents, according to Caldwell, is that in most societies parenting is now a low status job, so that parents do not approach it with commitment or pleasurable anticipation. (Section 3.25)

Parent intervention (programmes). Among the relatively small number of parent intervention programmes there have been some important initiatives. In the U.S.A. Skeels showed that after a small group of orphaned children had been individually 'parented' by mentally retarded hostel women for 18 months, they were sufficiently advanced to be adopted; 30 years later they had virtually all made a success of their lives, compared to a control group who had not had the early caring and who were all still a heavy burden on the State. Another American, Gordon, used trained non-professionals to establish a successful home visit and home learning centre programme. Other workers in that country have set up training sessions for mothers at pre-school centres, or used 'skilled graduate toy demonstrators' to visit the homes and work with children. Yet other workers have established programmes in which parents of infant school children have been encouraged to provide a more educational environment in the home. In the United Kingdom various programmes have likewise provided either group meetings or home visits. One programme aimed to help mothers identify the educational opportunities available in the home, rather than offering a prescribed list of activities. This was the E.P.A. Red House programme, which used educational visitors to work with children and parents in the homes. A large-scale Scottish programme required visitors to work with two and three-year-olds in the home, in front of the parents, for an hour a week. Yet other forms of parent intervention include the nation-wide playgroup movement - a parent-run initiative with thousands of groups in existence throughout the United Kingdom - and the American Portage programme. The latter is a highly structured programme which

was created to help parents to stimulate their handicapped children, but is now used increasingly, in different countries, in programmes aimed at disadvantaged parents. The methodological and conceptual limitations of many of these programmes are discussed, in particular the undesirability of working with children at home in front of their mothers, reinforcing the inadequacy that mothers might already feel. Many of the programmes have not been adequately evaluated and a great deal has yet to be learned about their effectiveness. (Section 3.25)

Goal-setting and strategies (action research). Halsey describes the concept of action research in terms of small-scale interventions, usually within administrative systems, accompanied by process and other evaluation but not bound by the constraints of empirical research; in this situation monitoring may indicate the need to change the original research goals. While Halsey rejects the 'neutral' evaluative role as limiting the participatory value of this type of research, it is possible to argue here that a rejection of too many of the tenets of formal and rigorous research can lead to a rejection of every firm methodological foundation in a research programme. Many action research programmes can be faulted for the limited nature of their data-gathering and analyses. Action research rightly implies flexibility and openness to change, but the more that the principles of rigorous evaluation are abandoned the more likely it is that the results will be rejected by the wider research community. (Section 3.31)

Goal-setting and strategies (the best focus?). Among the problems which still await resolution are the question of whether early intervention is more effective than later intervention. The limited results with pre-school institutional intervention do not provide clear answers. An even more fundamental issue is the question of whether parent or institutional intervention is likely to have better long-term effects. Although most existing research points to parent intervention as more effective, many psychologists and teachers appear uneasy about the prospect of entering the home and about the cultural implications of such a move. Yet there are other psychologists, such as Bronfenbrenner, Hunt and Burton White, who see the great potential of offering the parent guidance in the home. (Section 3.31)

Design considerations. While the conceptual assumptions and theoretical viewpoints of research workers tend to influence the design and statistical interpretation of research studies, it can be argued that the wider and more comprehensive the design, data collection and analyses, the less one-sided are likely to be the results and presentation. In reality a great deal of research design is timid and narrow, with data being gathered from only a few sources though often gathered with considerable accuracy - a procedure which heightens the reliability of the particular measures, but leaves little validity in the model as a whole. An added limitation is that many analyses tend to present only bivariate

relationships, possibly controlling for one or two other variables such as social class; yet in nature most situations are multivariate and the design and analyses need to reflect this reality. (Section 3.32)

Sample considerations. Problems in obtaining representative samples and in finding comparable intervention and control samples - particularly in a field situation where intervention parents are asked to volunteer for participation - are among the major problems which arise in intervention research. One method of coping with the problem of finding comparable intervention and control samples is to set up a number of different intervention groups for which programmes of differing types or lengths are provided, with only one or two control groups. Another difficulty which arises in many studies, even some of those designed by leading workers in this field, is the smallness of the samples. In statistical terms the small samples are particularly open to chance effects, so that only unduly large treatment effects are likely to be recognised as 'successful'. (Section 3.33)

Research perspectives. Various research workers have noted particular methodological flaws in pre-school intervention programmes. Miller and Dyer refer to the absence of systematic observations on whether the intervenors correctly apply the intervention curricula which are under scrutiny. Ambron, in a wide-ranging review of these programmes, finds a considerable gap between practice and the theory or research on which the practice is ostensibly based. Piaget's theories are frequently cited, for example, and yet he did not offer a theory of intervention; likewise Bernstein's theories do not necessarily provide the proper basis for a language programme. Smilansky, who undertook a review of pre-school intervention programmes for the World Bank, concludes that pre-school is not the 'cure-all' it was once thought to be; the evaluation of many programmes is flawed and the designs are primitive. However Smilansky himself bases his judgement of effectiveness almost entirely on the criteria of cognitive growth, ignoring the evidence on long-lasting academic improvements resulting from the more highly structured pre-school programmes. All three authors are in agreement, however, that programmes which focus on the parents are more effective than those which aim to teach the child directly. (Section 3.34)

Evaluation. Research without evaluation is an incestuous activity, since the insights it offers are only for the eyes of the participants or for those who are willing to accept the assurances of the participants about what has been achieved. An equal danger lies in the presentation of research findings in the form of the collected impressions of the researchers or of the participants themselves, rather than basing findings in the first place on the analysis of the evaluation data. By its nature, intervention research is applied rather



than 'pure', although a variety of fundamental theories can be tested in the course of the applied studies. Intervention research is also in essence of principal concern to administrators and policy-makers. In these circumstances it is surprising that the main and often the only criterion of success is whether a change has been brought about in the Intelligence Quotient or I.Q. of the children, despite the fact that most educational or psycho-educational intervention studies are focused ultimately on an improvement in academic attainment. While the predictive power of I.Q. - itself a highly culture bound and culture normed indicator - in predicting later academic performance is advanced as a reason for this bias, early academic attainment has an equal or greater predictive strength. The human environment surrounding the child has an even more powerful predictive value for later performance, and assessments of changes in this environment may well yield more reliable evidence on the success or otherwise of intervention programmes than do I.Q. or academic attainment scores on the child. Yet another area of increasing importance in evaluation is the examination of the 'process' which occurs in the interaction between intervenor and intervened, helping to determine how the programme is interpreted by those to whom it is offered. (Section 3.41)

Cost-benefit analysis. Cost-benefit analyses would be likely to show that highly successful studies such as the Milwaukee Project are completely impractical because of their vast cost, and are therefore of little or no applicability to a population. It is seldom however that such analyses are undertaken, despite their importance in showing the usefulness or otherwise of the intervention concepts for public policy. One of the few exceptions was the detailed and important study undertaken by Weber and colleagues in the Weikart (Ypsilanti) pre-school project; this showed a reasonable rate of return for the investment in the intervention. Another analysis, carried out on behalf of the Brookings Institution, showed however that because of the limited evaluations and other shortcomings, the few pre-school compensatory programmes which could be assessed (Weikart's was not then available) appeared to be more expensive than compensatory programmes undertaken in the school years. (Section 3.42)

Dissemination and public policy. In the dissemination of the results of intervention research, political and administrative policies inevitably become the object of attention. This in turn can lead to attempts to 'politicise' the research itself, or to attempts (sometimes by politically committed academics themselves) to question the findings of others if their acceptance implies the adoption of policies contrary to those favoured by the academics. Both the Head Start and Follow Through intervention programmes in the U.S.A. have been the subject of heated academic and political debate over many years, while in Britain the Bennett (teaching styles) and Rutter (school ethos) studies have aroused con-

siderable discussion at the same two levels. Bronfenbrenner considers that rather than believing that social policy needs science, it is science which needs social policy in order to provide it with vitality and validity, and particularly in order to ensure that research becomes ecologically valid, taking into account the great many other factors which it has tended to ignore. Another important consideration is the value system or 'hidden curriculum' of the researcher her or himself, and the extent to which research results and proposed remedies reflect the needs of the middle class child (and/or parent) rather than those of the disadvantaged 40 per cent of the population. (Section 3.43)

The present study: concepts and hypotheses. The concepts underlying the study are described; the aim is to build on the findings from the review of the research on early attainment and intervention, and to explore a new form of parent intervention within the British pre-school context. Three hypotheses for the study are formulated and presented here, for testing against the data to be collected in the study. (Section 3.50) The hypotheses themselves appear in full later in this synopsis, in section 7.30.

#### From Chapter 4: Intervention Project

Sample population characteristics. The Metropolitan Borough where the study was undertaken is mainly a working class area, with 70 per cent of the population being categorised in social classes 3B, 4 and 5; only 20 per cent of the population are in owner-occupied housing, with most of the remainder being Council and private tenants. At the time of the study some 25 per cent of the population had been born abroad, more than half of these being newcomers originating in the West Indies. Unemployment in the area is well above the national average, and for young Blacks in particular it is even higher. Some 17 per cent of families are single parent, and between 20 and 25 per cent of mothers are in employment. (Section 4.221)

Nursery facilities and distribution of under-fives. The majority of three and four-year-olds in the Borough are in some form of pre-school provision, with over one-third attending nursery classes, nursery schools or having entered reception class as 'rising fives'. It is estimated that only 10 per cent of the four-year-olds are at home with their mothers or with unregistered child-minders. (Section 4.222)

Provision of books. Libraries are thinly spread in the densely populated areas of the Borough, and bookshops appear only within the largest shopping complexes, well beyond walking distance for most residents. There is an almost total absence of books reflecting the West Indian and Asian cultures in the book-

shops, although such books can be borrowed from the Borough's libraries. (Section 4.223)

Selection of sample schools and their children. The criteria used in the random selection process, and the reason for certain exclusions from the sample, are set out in detail. (Section 4.224)

School, class and staff characteristics. Each of the schools in the sample was committed to the major goals of fostering language and achieving literacy for its reception class entrants, and a variety of methods were used to reach these goals. The philosophies of the schools' Nursery classes ranged from a fairly formal to a fairly progressive approach. All the schools were particularly concerned to involve the parents in educational and social activities, although this was not attainable with all parents. However for most Nursery parents there was regular contact with what might be termed the educational system, in the form of daily meetings with Nursery staff when the infants were brought to or taken from the Nursery class session in the morning or afternoon (most children attended only half day sessions). At these informal meetings with staff members a variety of problems concerning not only the children's welfare and progress but also those of the mothers (or occasionally fathers) and of their families were discussed; this occurred particularly when one of the staff members was herself older than the mothers and had come to be looked upon as a counsellor. (Section 4.225)

Selection of parents. The reasons for dividing the sample firstly into working and non-working parents ('working' referring to paid work outside the home), and then randomly dividing the non-working parents into reading and mathematics groups, are set out in some detail. The reasons why the parents volunteering to participate in the programme were not divided further into intervention and control groups are discussed here and in later sections. (Section 4.226)

Method. The design and implementation of the study can be defined methodologically in terms of three differing forms of evaluation and a set of 'treatment' programmes. The forms of evaluation include an assessment of a number of the educational attainments of the children both at the start and end of the research phase, a period extending over 22 months; an assessment of a variety of the cognitive and meta-cognitive skills of the children, both at the start and half way through the research phase; and an assessment of a variety of characteristics of the home environment, as derived from an interview with one or both parents. The 'treatment' programmes consisted of the fortnightly (or, in a few cases, extended six-weekly) reading and mathematics programmes given to the 25 parent groups, with a quantitative assessment of the attendance of each parent at these programmes, yielding scores for use in the analyses. The

choice of the various assessment instruments, and how they were selected or developed and piloted, are set out in this and subsequent sections. A similar extended description of the parent programmes is presented. (Sections 4.30 et seq.)

### F i e l d   W o r k

The intervention programme devised for this study has been based on the evidence, referred to in the earlier review of the research, that intervention is more likely to succeed with parents than with children, since change in the parents is likely to lead to a permanent change in the child's environment. Various findings suggested that the intervention should be in the form of a semi-structured programme aimed at persuading parents to examine and if necessary to modify their child-educating behaviours, rather than attempting to change the attitudes or beliefs of the parents. Its ultimate goal has been to assess whether programmes such as that tested in this study are sufficiently effective to be put forward as an example of what could be provided by teachers for all parents in disadvantaged areas during the year or more when their children are attending nursery classes, nursery schools or other forms of pre-school provision.

Because of the limitations of time and resources, it had to be recognised that any intervention programme could only be moderate or minimal rather than optimal. Where it seemed justified the strict tenets of a formal experiment have been modified, so that the study is close to what has been defined as action research, but without sacrificing rigour in the evaluation itself.

The field work which has been carried out for this study, with over 200 parents and their children and with the cooperation of staff at eight schools, can be summarised as follows:

- September 1976: Discussions with Metropolitan Education Authority; liaison with heads and nursery staff at six randomly selected infant schools (five in disadvantaged areas and one in an advantaged area) and at two further 'pilot' schools, also in disadvantaged areas.
- October 1976: Preparation of pilot programmes and assembly and preparation of evaluation instruments, including child tests and parent interview protocols; piloting and modification of evaluation instruments in first pilot school; piloting and modification of interview protocol with parents from second pilot school; reliability tests on Pre-Test (Nursery) child assessment battery, at first pilot school.
- November 1976 to March 1977: Nursery testing of all the Nursery class children

(within the age range 3:9 upwards) at the six principal study schools; interviews with parents in their homes during the same period; administration of pilot programmes in reading and mathematics to two parent groups at second pilot school.

April to July 1977: Administration of reading and mathematics programmes to 25 parent groups at the six sample schools; pilot and reliability tests on the Mid-Test battery, at first pilot school.

April 1977 to April 1978: Administration of Mid-Test battery to sample children soon after each child reached Reception class (with testing in the nursery of the very few who were still there at the end of the Mid-Test period).

March 1978: Pilot tests on the Post-Test battery, at first pilot school.

May-June 1978: Administration of Post-Test battery to all surviving sample children (159), either in their six original schools or in 20 other schools to which members of the surviving sample had moved. Reliability tests on the Post-Test battery with subsamples at the study schools.

1978 and part of 1979: Development of V-ridge regression computer programme and principal path analyses of the data.

In total, data were collected on approximately 50 continuous variables and 30 categorical variables, over the three phases of the study, for each of the surviving 159 parent-child dyads in the sample. Data collected before the attrition losses were used to compare attrition and surviving samples. Computing advice and direct line access to the University of London Computer Centre were generously provided by the Department of Mathematics, Statistics and Computing at the Institute of Education.

#### From Chapter 5: Analysis

Statistical options. A great deal of statistical and methodological thinking suggests that data analysis requires a comprehensive and multivariate approach rather than the simpler options which are often used when examining research findings. For example, many important studies in field research situations have relied on t-tests and similar unsophisticated criteria when reporting results. It is less often that analysis of variance, regression, path analysis, discriminant analysis and similar techniques are used to draw a much fuller picture of what has been found. Namboodiri and colleagues have shown that complexity in analysis can in fact help to overcome the limitations in the quality of the data, so that the argument that the quality of the data does not 'justify'

more advanced analyses is in fact counter to reality. A variety of statistical options are examined here. (Sections 5.10 to 5.103)

Bivariate relationships. It is shown that bivariate relationships have only a restricted usefulness on their own. Tukey, for example, regards the correlation coefficient as only tangential to data analysis. Others have shown the danger of relying on correlations to define causation. While bivariate relationships have some value as preliminary pointers, their main function is to serve as building blocks for multivariate analyses. (Section 5.11)

The multivariate approach. The merits and limitations of multivariate models of analysis are examined. Among the problems are the question of how comprehensive and complete is a model, since no limited set of variables can fully define a relationship in the social sciences. Likewise a selection of the set of variables thought to define a relationship most adequately may differ somewhat from the set selected by another analyst on other theoretical grounds. The 'sensitivity' of a set of variables in fitting a model is a related issue, since a less sensitive set would be more likely to yield a chance fit to the data. A further problem in multivariate models can arise from the phenomenon of reciprocal causation, since it is not always certain as to how much particular environmental factors, personal characteristics and personal behaviours interact with or 'contribute' to each other; the assessment of directionality and the taking into account of the bi-directionality of influence can considerably complicate the analysis of a model. (Section 5.12)

The interpretation of variance. A great deal of statistical analysis relies on the interpretation of the relationships between the variances of the variables within a model. Two powerful methods of analysis within this paradigm are analysis of variance and regression analysis. They each have their strengths and weaknesses. While analysis of variance offers strong evidence on the differences between samples and the strength of these differences, taking into account a considerable number of competing variables, regression analysis offers a more direct interpretation of the predictive powers of variables within a relationship of multiple and competing predictors. Another contrast lies in the fact that analysis of variance is based on the interpretation of relationships across cells or groups of cases whose existence (and similarity) is defined by a combination of categorical values and specified ranges of values within continuous variables; regression, in contrast, attaches to each individual his or her own unique set of categories and values of the variables within the model, although this in turn contributes to other statistical problems. For reasons set out in detail in the relevant section, regression analysis was chosen for the present study. (Section 5.122)

Causal models. There has long been a major divide between what might loosely be termed the 'English' and the 'Continental' approaches to the problems of causality.\* For centuries many English philosophers have hesitated to concede what appears to be the absoluteness of defining a causal relationship, preferring to offer the more cautious interpretation of an association; this is linked in turn to an inductive approach to science and with it the acceptance or rejection of hypotheses about associations. In contrast, Continental philosophers and a minority of English philosophers have seen analysis as the development of deductive causal models within which particular hypotheses can be examined. While probability criteria enable firm judgements to be offered about the strength of associations, deductive causal models enable a more interesting if also more tentative explanation to be offered for a set of relationships. (Section 5.13)

\* Richard Pring discusses the question of causality in his contribution to Wragg et al (1980). As a philosopher, Pring recognises the conceptual difficulties in regard to causation. But the fact that there are a multiplicity of possible causes of a particular result simply shows that one can never be certain of the claimed causal relationship, in the same way that there can never be certainty about other conclusions in life. Likewise the objections that there are 'subjective meanings' which may affect a person's behaviour, rather than the meanings or 'causes' ascribed to the behaviour by an outsider, do not necessarily disprove the concept of cause as an explanation or 'reason why'. Even if it were possible to determine the subjective meanings ascribed to situations and behaviours by an individual, there would still be room for outside explanations since a person's ideas and concepts have themselves been acquired from elsewhere, for example, through participation in public life. According to Pring "there is no reason why one should rule out completely reference to causal factors.... to explain some aspects of behaviour".

Reliability and disattenuation. The reliability of variables is sometimes wrongly seen to be of more importance than their validity, or the validity of the model in which the variables' relationships are being examined. In fact validity is a far more fundamental concept. Variables with a reliability as low as 0.5 can still play a useful part in a model, provided both the model and its variables are reasonably valid interpretations of a situation. On the other hand a low or dubious validity for the variables or for the model as a whole can call into question both the research hypothesis and the manner in which it is being explored, even if the variables in question are being measured with a reliability close to unity. Reliability is however useful as an indication of the competence or accuracy with which variables have been measured or evaluated. Indices of reliability have been determined for all the variables used in the analyses in the present study; these indices are employed in particular to disattenuate the matrices of correlations which are employed as the basis of subse-

quent multiple regression relationships. (Section 5.14)

Validity assessment. Particular attention is given here to an assessment of the validity of the many variables used in the study and of the wider validity of the multivariate models which are finally explored. A possibly important development in this study has been the creation of several new measures of validity in response to Cranbach and Meehl's call for a wider definition of construct validity - validity within the 'nomological net' of an interlocking system of laws; as these authors point out, it is the networks from which the constructs derive their meaning. This wider concept of validity has been given statistical definition in the present study, either in the form of what is here termed the nomological validity of single variables within a given network of multivariate relationships, or in terms of a more comprehensive assessment of the nomological validity of the multivariate model itself. The new measures of nomological validity have been assessed for all the variables used within each major analysis, and the more advanced measures of the nomological validity of these major models of relationships have been determined. By definition these two aspects of validity are intimately related to each other and in combination they provide indices by which the usefulness of particular variables and particular models can be assessed. To the extent that these measures provide quantitative indices for defining the validity of a multivariate model as a whole, as well as of its constituent variables, they may be of some use in the methodology of both the psychological and educational sciences. However the indices are not intended to serve as definitive criteria of the acceptability or otherwise of variables or models, but rather as measures of how usefully the variables and the model itself contribute to the analyses within a particular totality of relationships. (Sections 5.15 et seq.)

The significance concept. The many limitations of the concept of significance and its customary application are reviewed. Whereas significance levels should be seen simply as an indication or measure of the improbability of a finding under the null hypothesis - i.e., the hypothesis that there is no effect, difference or other hoped for result - they are all too often used as 'proof' of the strength of an obtained result. In this study significance levels will be used only as cautionary indicators rather than as determining criteria for the strength of any findings. Thus variables will not necessarily be excluded from path models on the grounds that the probabilities of their derived coefficients are modestly above the conventional significance level of 0.05, if parameters such as unique variance predictions or other considerations suggest their inclusion. The very few breaches of this significance convention in the study are noted when they occur. (Section 5.16)



Age and time concepts. The effect of 'controlling for age' in a multivariate model often serves to remove part of the variance of an outcome or dependent variable which should rightfully be shared with other predictors such as length of experience in a classroom. Age is therefore treated here as an equal rather than a prior predictor in the analyses, the child's age at each of the three stages of assessment being entered alongside other predictors at these stages. The time concept has proved to be more complex than age has in the analyses. Although closely related to age, the two variables are not interchangeable. The rationale is presented for creating two time variables, Time in Nursery and Time in Reception, reflecting the time spent by each child in the nursery or reception class between the first and second or the second and third testing sessions, after adjusting for the child's 'cohort age' and the distribution of ages in the child's own school. The possibility that the variances of the two time variables can predict part of the outcome variance of academic attainment is explored in subsequent analyses. Other time-related variables which are to be used in the analyses include measures of the number of programme meetings attended by the parents. (Sections 5.17 et seq.)

Multiple regression. The theory and application of the multiple regression technique are discussed at some length, in view of the importance of this technique in the planned analyses of the study data. Three particular problems are reviewed. The least squares capitalisation on error in calculating  $R^2$  is controlled by the application of a correction formula. The serious objections to the use of automatic stepwise algorithms are set out and a formal decision is taken to rely entirely on simultaneous multiple regression, followed by the selective removal of poor predictors and a further regression on the remaining predictors. Finally, the arguments for attributing the concept of 'suppressor variables' to these variables whose regression coefficients are in the reverse direction to their correlations with the dependent variable are seriously questioned, on the grounds that it is the shortcomings of the least squares algorithm which are responsible for the aberrant behaviour. The use of cross-validation techniques is described and proposed as a method of checking on the reliability of the coefficients to be determined in the proposed analyses. (Sections 5.20 et seq.)

The limitations of Ordinary Least Squares. The most important limitation of the multiple regression technique, when using the ordinary least squares algorithm, is the unreliability of regression coefficients derived from small samples or from samples in which high levels of multicollinearity exist; this problem is examined in depth. (Sections 5.21 et seq.)

V-ridge regression: theory and evidence. The rationale and theory of ridge regression, a recently developed technique aimed at overcoming or at least mini-

missing the unreliability of regression coefficients, are described in some detail. The technique is a relatively new one. In the present study a new non-stochastic method of ridge regression is developed on the basis of an algorithm pioneered but not developed further by the American statistician Vinod. While most methods of ridge regression are stochastic and therefore require subjective judgement for the determination of what is termed 'ridge k', the present method is entirely automatic in its determination of the ridge constant, so that its results are objective and repeatable. (The selection of variables for inclusion or exclusion does of course remain in the hands of the analyst, based on a judgement of the results of each regression equation.) Cross-validation has shown that the derived coefficients of what has been named here as 'V-ridge regression' are nearly always more stable than those of least squares regression, and remain relatively unchanged with changes in the number of variables entered into an equation (i.e. in a situation where poorer predictors are eliminated). Extensive evidence is presented on the effectiveness of this technique. Its development is seen as one of the most important and potentially far-reaching innovations of the present study. Many of the study's findings and some of the results supportive of findings from other studies would have been impossible without the power, reliability and added sensitivity of the new regression instrument described here. (Sections 5.22 et seq.)

Path analysis. The particular value of path analysis, as a method for presenting a model of multivariate relationships over time, is described. Among its limitations are the reliability of the coefficients derived in the multiple regression equations which link the variables, and the difficulty of ensuring correct model specification - in other words, how justifiable and how complete is the set of variables included in a particular model. While the use of V-ridge regression reduces the problems of variable unreliability, uncertainty about model specification will always arise with hypotheses about a complex set of relationships. The particular form of path analysis planned for this study is noted. (Sections 5.30 et seq.)

From chapter 6: Results

Child tests. The testing programmes are described. Sample children who left the study schools were traced to more than 20 non-sample schools to carry out the post-tests. It is shown here that there were slight to moderate differences between the attrition sample (those children who could not be traced) and the surviving sample; 159 were traced for the post-tests, out of an initial 204 parent-child dyads. Out of ten cognitive and attainment variables selected for the comparison, six favoured the surviving sample and four the attrition sample.

In five of the six tests favouring the survivors, t-tests of the differences showed a probability below 0.05. The top and bottom one-sixths of the attrition and surviving samples, on seven of these variables, showed no particular tendency for higher or lower performers to leave the area. (Section 6.10)

Parent interviews. The administration of the parent interviews in the homes is described. It was found that parents raised many issues of educational concern during the interviews, particularly in regard to their worries over the schooling of their older children. A majority of the parents expressed the desire - without being asked - to see someone from the school for a home visit at intervals, yearly or more frequently. In the multi-cultural situation a variety of other issues relevant to the concerns of different ethnic groups were raised by the parents. (Sections 6.20, 6.21, 6.22)

Goals and ethical issues. Many parents were concerned about what they considered to be the undue restlessness or hyperactivity of their children, and asked for advice on this and other problems in regard to the education and rearing of their children. Their requests emphasised the conflict between the view that a neutral interviewer should not 'intervene' by offering advice on issues which might possibly be of some relevance to the study, and the ethical desirability of assisting parents who felt seriously burdened by specific educational problems. It was only in a very few cases that any advice was tendered, when there were major ethical imperatives which weighed more heavily than the formal neutrality of the research. Examples of these situations are cited. It is not considered that the results were influenced in any way. (Section 6.23)

Comments on nursery education. One of the interview questions asked parents to say what help they thought the nursery class had been to their children. A great many of the parents - the majority in the disadvantaged areas - used the term 'playschool' when referring to attendance at nursery. Figures obtained from a series of overlapping categories of answers showed that 60 per cent of parents thought that socialisation or social caring were particular functions of the nursery, these functions being described by terms such as 'teaching manners', 'teaching discipline', or 'calming'. Approximately 27 per cent mentioned the goal of language development. A total of 46 per cent spoke of general development, such as creative activities. Only 13 per cent saw the nursery's function as educational, with the highest proportion coming from one area where the nursery class offered number games and some other broadly educational activities. Alongside a considerable appreciation of the general nursery goals there was some criticism of what parents considered to be the failure of the nursery classes to provide an educational experience for the children. (Section 6.24)

Reading activities (responses). There were many interesting responses to the

questions aimed at assessing how frequently parents read to their children. The range of answers varied from a total unawareness that the parents had any role in this matter, to an acute awareness of the responsibility of the parents. Some parents said they would welcome guidance from nursery teachers as to how they should teach their (pre-school) children to read. The biggest obstacle seen by many disadvantaged parents (to the idea that they might initiate early reading in their children) was their concern as to whether they would be using what they termed the 'right method', and whether the parents' methods might clash with those of the schools, resulting in what many parents thought might be penalisation of the child or damage to its early education. (Section 6.25)

Reading activities (library usage). The parents' answers showed an extremely low level of library usage in all the disadvantaged areas. Even positive references to 'library books' often meant only those books which older siblings obtained from their school libraries. The overriding concern, when the question of borrowing books from public libraries was mentioned to parents, was fear of the damage that the children might do to these books. The small sample of advantaged parents were aware of the value of library books and were usually members of libraries. (Section 6.25)

Mathematics activities. Parents found it difficult to answer questions on the practice or otherwise of number-oriented activities and games involving spatial awareness, even when these were described simply. They were able to recognise that reading to children, or fostering reading by the children, was something that might or might not occur in the home, but they were often puzzled by the interview references to numerical or spatial activities. Counting steps up and down the blocks of high-rise flats was however mentioned by many parents. (Section 6.26)

Attitude to school and school work. In general, schools were seen by most parents as having the main responsibility for preparing their children for work in the classroom. Many parents also thought that schools should inculcate discipline, whatever the state of discipline in the home. (Section 6.27)

Television viewing. Television viewing, whether quantified in terms of viewing hours or measured by an estimate of the degree of parental control over their children's viewing, bore only a minimal relation to the children's cognitive levels or academic attainments. Viewing hours also bore little relation to the social environment of the home. For some parents television was seen as a kind of pacifier for the children; others saw it as an educational experience. The mean weekly viewing time of the whole sample of 3:9 to 4:9 year old children was 19.2 hours. For the advantaged children in the sample viewing time was 17.9 hours. These figures are comparable with British and American viewing

data for young children. (Section 6.28)

An ideal parent? Based on activities cited by different parents in the interviews, a composite 'ideal parent', in relation to the goal of an early introduction of reading and number activities for the child, could be seen as someone who provided an environment which involved: one or two hours direct interaction with the child each day; the conscious organisation of the mother's working day (whether working at home or outside) to leave some time for the child; the playing of 'I spy' or other language games with household objects; the utilisation of shopping and bus journeys to increase the child's general awareness and possibly to point out bus numbers and some shop names; the playing of finger games to develop number awareness; the encouragement of the matching of domestic objects; the facilitation of water play; the guidance of the child in the choice of television programmes; the purchase of educational toys; the introduction of further reading and number activities as the child's interest develops; and not waiting for the child to take the initiative but actively fostering its interest in educational activities. (Section 6.29)

Parent programmes. Liaison with the sample mothers and fathers over the nine months covering the interview and programme period showed that a wide range of life problems, often serious, faced these parents during the time that their children were in the nursery class. Many of the worst problems were concerned with the effects of disadvantage. As life was often lived at a day to day level of worry over their immediate problems and needs, it proved particularly difficult for many parents to remember the meeting days in advance, even with the variety of methods used for reminding<sup>d</sup> them. (Sections 6.30, 6.31)

Attendance pattern. Out of the final surviving sample of 159 parents, 99 attended one or more of the reading or mathematics programme meetings, 26 agreed to attend but did not attend any meetings, 31 were working parents who by definition could not attend, and 3 refused the invitation to attend the programme (at the time of the interview). Approximately one-quarter of the disadvantaged parents worked, compared with only one-tenth of the advantaged sample. Of the total of 99 parents, 51 were still attending at the seventh and/or eighth (last) meeting of their particular groups. This compares well with adult education programme attendance patterns. At the end of the programme some groups still contained four or five parents, while others had been reduced to the one parent who was keen enough to continue attending. (Section 6.32)

Inherent problems (undue pressure by parents?) Among the problems which arose during the programme was that of persuading parents who themselves read little and had only limited educational awareness, to take an interest in the early educational stimulation of their children. Parents also lacked self-confidence

and did not initially believe that their children could be enticed to take an interest in reading or in spatial and other mathematical activities. One issue which concerned teachers in particular was what they saw as the danger of undue pressure by parents on their children. Yet only 2 out of nearly 100 parents taking part in the programme showed any tendency towards undue pressure. In both cases the pressure was already being exerted prior to the start of the programme. In one case the parent (a father) was easily persuaded to alter his style of demanding totally correct responses. The other case concerned an ethnic minority woman whose menfolk had long been pressurising the child and trying to force it to learn its letters and numbers. The mother was given advice on dealing with this situation tactfully, and later the child's teacher reported a remarkable change in the child, from sullen apathy to lively interest. In the overwhelming majority of cases the situation was one of modest interest or of unawareness and neglect of the educational potential of the child, rather than one of the parents pressing for achievement. If the evidence of both the advantaged and disadvantaged groups in this sample is typical, the 'danger of undue pressure by parents' is largely an educational myth, supported only by isolated incidents but serving to reinforce the gap between the professional and the non-professional. (Section 6.34)

Inherent problems (individualisation) It became apparent that the most fundamental difficulty with the group programmes was that they could not be geared to the individual needs of parents. By their very nature such programmes are either suited more to the competent and successful parents, or operate at a much simpler level to suit those who have difficulty in following the programmes; more often it is the successful parents who do much of the talking at meetings and eventually those parents who are struggling with the programme stop attending meetings. Only a home-oriented programme, which was not possible in this study, could meet the diverse needs of each parent. (Section 6.34)

Parents' views on the programmes. All the parent groups were asked to give their views on the programmes at the end of the last meeting of each group. Among the points raised was the question of whether children who had gained additional educational skills as a result of the programme would become bored during their initial period in school, when teachers might be concentrating on fostering skills (in the rest of the class) which the child had already gained; in reply it was suggested that competent teachers were able to gear their teaching to the needs and levels of individual children. Various parents stated that the value of the programmes was that they provided what parents had been wanting to do, or in some cases had already started doing, for their children. The parents valued the opportunity to interact with other parents in this context, although it was recognised that some of the more shy or less 'competent' parents tended to

opt out of the groups. Many parents expressed the wish that the programmes could have been adapted more to individual children's needs; the difficulty of achieving this in group programmes was discussed. Parents noted the problem of coping with the changeability of their children - sometimes they were keen to play the educational games, and sometimes they were not. When asked about the preferred frequency of meetings, there was no consensus as to the most desirable intervals; a large minority favoured monthly meetings, on the grounds that the two-weekly programmes were difficult to attend given their other commitments in running their homes. Many parents appreciated the fact that the themes suggested at each meeting could be tackled at different levels, according to each child's stage of development. The structuring of the programmes and the demonstration of more advanced mathematical activities was also welcomed. It needs to be recognised that the views of the minority of parents who opted out of the programmes at an early stage are not represented in this review. (Section 6.36)

Lessons for future programmes? A retrospective examination of the administration and content of the programmes suggested a variety of possible innovations for other programmes of this nature, including: written notification to parents, only a few days prior to each regular meeting; a partial individualisation of the programme, with a smaller part of each (monthly two-hour) meeting devoted to private discussions with individual parents; structured procedures for bringing parents up to date on the work of meetings which they have missed; greater efforts to persuade parents to join and use the local libraries; more repetition of activities with children, to reinforce newly acquired skills; the possible recruitment of 'parent group leaders' to undertake the organisation of meetings and to motivate other parents to attend; the combination of reading and mathematics programmes within a single programme; and the provision of this kind of programme over seven monthly meetings for all nursery parents, each meeting occupying an entire morning or afternoon in a venue close to the nursery class. (Section 6.38)

Broad characteristics of the data. Out of 81 variables derived from the various assessments and other measures in the study, a total of 52 continuous and 8 categorical variables were selected for use in the multivariate analyses. The distributions and other major parameters of each variable were determined. (Section 6.41)

Sub-sample characteristics. Among the findings of interest - some expected and some unexpected - were the following:

- A considerable difference in the performance of children attending morning and afternoon nursery class sessions, with the cognitive and academic attainments of morning children being generally higher. Possible reasons for this

difference were explored.

- Girls showed the expected superiority to boys on all items except the cognitive (spatial) picture completion test.
- There were very considerable differences between the nursery children at the five schools in disadvantaged areas, and the children from the advantaged school; the latter children performed at levels up to 40 per cent above those of the disadvantaged children. In terms of educational policy and governmental resource allocation, data such as these indicate the extent of the differences that exist prior to the start of formal schooling; it suggests that much of what is claimed to be inadequate performance by inner urban schools cannot be laid at the doors of the schools, when the intake itself is so totally different between advantaged and disadvantaged areas. It can be noted that the sample was reasonably representative of the area's children as a whole and did not represent isolated pockets of deprivation.
- In both the interview variables and attendance at meetings, advantaged parents achieved higher scores than those of disadvantaged parents.
- There were virtually no differences in the performance of Black and Non-Black children on 7 key performance variables assessed in the first (pre-test) battery; on an eighth variable - rhythmic tapping, a test of the ability to copy increasingly complex patterns of taps - the Blacks were well ahead, having nearly double the scores of the rest of the sample at the pre-test, and scoring 50 per cent ahead at the mid-test. This supports the finding of a more wide-ranging cross-national study (Igaga and Versey, 1978), showing that rhythmic ability and synchronisation scores were superior in a Ugandan child sample, compared to an English child sample.
- In contrast with the general equality of cognitive performance across the two groups, there were differences in the parent interview scores for these groups. Both the reading and mathematical environments in the Black (and mixed race) homes were somewhat lower than those in the Non-Black homes (mainly Whites and Asians), as measured by the interview variables; the mean attendance levels at the programme meetings showed larger differences in the same direction. These parental differences may reflect an interview protocol designed in terms of European cultural pre-conceptions as to what is academically desirable in a home environment; the protocol did not contain measures to assess the social and emotional warmth and security provided by the parents for their children - for the latter criteria, observation suggested that most Black parents would have scored highly, probably well ahead of most Non-Black parents. It also needs to be emphasised that on almost every variable some Black homes were found to be among the handful of top scorers for the sample as a whole, and equally there were some Non-Black homes whose scores were at or near the bottom of the



parent interview scales. The evidence taken as a whole however does suggest that it is the differing levels of practical academic awareness in the homes which contribute to these ethnic differences in the interview responses. (Section 6.42)

Comparison of parent group characteristics. The differences noted between the various programme groups include the following:

- Nearly all the variables favoured those who attended the programmes, compared with those who accepted the invitation to attend but did not come to any meetings.
- There were only random differences between those who accepted the invitation to attend the programmes and were then randomly allocated into reading groups, and those who accepted and were randomly allocated into mathematics groups; however nearly all the variables favoured those who attended the maths meetings, over those who attended the reading meetings. This suggests that it was only the more committed parents who attended the mathematics programme, since its goals were not as apparent or desired (for nursery age children) as were those of the reading programme.
- The variable scores for the Disadvantaged Working Group parents and children were higher than those for the Disadvantaged Non-Working Group (the latter consisting almost entirely of those who accepted the invitation to attend programme meetings). However the reverse finding occurred with the Advantaged parents and children, where the scores of the few dyads in the Advantaged Working Group were lower than those for the Advantaged Non-Working Group.
- As a further complication, it was found that the children of Working Group parents (in the sample consisting of all nursery class children between 3:9 and 4:9) tended to be older than the children of Non-Working Group parents, the difference between the mean age scores being about half a standard deviation. (Section 6.43 et seq.)

The experimental design and analysis. The difficulty of establishing a control group within the present study are discussed in detail. Apart from the need to have a large enough sample for the planned multivariate analyses - random division of the sample into programme and control groups would have halved the potential numbers - the various differences highlighted in the previous section (see above) and other differences between and within the schools and parent groups, indicated that direct comparisons between groups, on the basis of mean outcome scores of reading and/or mathematics, would not yield valid comparisons in view of the many complicating factors such as age and uptake differences between groups and sub-samples. This confirmed the original decision to rely entirely

on multivariate models for the principal analyses of each group of interest. (Section 6.44)

Reliability determination. Test re-test assessments of the reliability of a number of the variables used in the study yielded coefficients ranging from 0.47 for a simplified version of Kagan's Matching Familiar Figures, to 0.97 for the Information scale of the Wechsler Pre-School and Primary Scale of Intelligence, and 0.97 for the Infant Reading Test - a test developed in a previous study, with a 'reading equivalent' range from 4 to 6 years. Assessment of the reliability of parent interview variables was more problematical, since there are two aspects to reliability here: the reliability of the interviewer's scoring of the parental responses, and the reliability of the parental reports on their various behaviours; a combined index was calculated, based in part on other parental reliability data culled from the research literature on interviewing. Reliability coefficients were obtained for all the variables used in the multivariate analyses, enabling the application of disattenuation procedures. (Section 6.51)

Validity determination. The nomological validity was determined for all the variables used in each of the ten path models. (A brief reference to these newly developed indices appears earlier in this summary.) The redundancy index and other nomological validity measures for each of the multiple regression equations used in the ten path models was then determined, indicating the relative strengths or weaknesses of each model. (Section 6.52)

Nomological validities of individual variables. An examination of the nomological validity of individual variables shows particularly interesting characteristics for two variable groups:

- The variable with one of the lowest contributions to outcome variance over the ten models, and generally possessing the poorest nomological validity, is the English Picture Vocabulary Test, despite its good correlations with post-test attainment (ranging from 0.27 to 0.76), and its high measurement reliability as reported in the literature. The performance of this variable differs considerably between the disadvantaged and advantaged models. For all the disadvantaged path models the E.P.V.T. makes no independent contribution when entered alongside the cognitive and attainment variable groups. For the advantaged path model, however, E.P.V.T. makes a meaningful contribution at each of the three stages of regression analysis, in competition with the cognitive, ability and parent environment variables. This is a most puzzling finding, particularly as the raw correlations with the outcome variable are around 0.50 for most of the disadvantaged groups. Such high correlations suggest reasonable measurement reliability. The difference in relative predictive power (and thus in nomological validity) between the advantaged and disadvantaged groups suggests that

E.P.V.T. is largely duplicated by other variables in the latter models, but not in the advantaged model. It also suggests that for disadvantaged children, vocabulary skills as measured by the E.P.V.T. play a less prominent part in academic development in the early years than they do for advantaged children, despite the well known long-term importance of language development for all school achievement.

• The nomological validities of the two main 'time' variables, Time in Nursery and Time in Reception (adjusted to take account of the child's 'cohort age' and standardised for each school), shows the reverse phenomenon to that outlined above. Here these variables contribute modestly to most of the disadvantaged path models, but not at all to the advantaged model. What may be occurring is that the contribution of the parents' home environment, in academic terms, is so powerful in the advantaged sample that it totally outweighs any academic contribution from the school in the nursery and for part of the first year of school itself; for the advantaged child, who will usually have had a fairly academically oriented upbringing, with frequent book reading, outings to places of some intellectual interest, and the provision of numerous educational toys and games, the value of the nursery and first year school experiences may be principally that of learning to mix socially and to adapt to a large group situation in which the child is no longer one of the few or only centres of adult attention. (Section 6.521)

Nomological validity parameters for regression models. The regression equations which are used to create the latent variables are examined for their nomological validity, as assessed by the newly developed redundancy index. In general most of the basic regression equations (in which the outcome variables are regressed on to groups of conceptually similar variables prior to creating a latent variable for each group of variables) show satisfactory redundancy indices varying from below 1.0 - highly satisfactory - to just over 2.0. However two variable groups in particular tend to yield unsatisfactory redundancy indices. The parent academic environment in the disadvantaged models ranges from 2.2 to 6.2, indicating only limited nomological validity; in contrast, this same parent variable group has a very sound (low) redundancy index of under 1.5 in the advantaged model. The second group showing only limited nomological validity are the variables comprising need for esteem and need for security at the nursery and reception levels. The only low redundancy index for these variables occurs in the path model for disadvantaged Black children, suggesting that for them the needs variables have a particular significance as (negative) contributors to attainment. (Section 6.522)

Data: age and time variables. In a discussion of the different measures of age and time - the latter being an assessment of the periods spent in the nur-

nursery and reception classes during the study - it is shown that the correlations of these variables with the variable scores for the whole sample contain a number of oddities which are not a reflection of any lack of reliability but rather a pointer to the disparate nature of the various groups within the study sample. Because of holiday breaks (which varied across schools) and other time and age factors, including the absence of some sample children through illness or family visits to other cities, differing ages of transfer from nursery to reception, and a bias towards a younger sample in the advantaged school, it is shown that reliance on crude correlational measures or simple 'corrections for age' would be highly misleading in analytical terms. The alternative of using these variables as independent predictors in each of the multivariate path models is justified. (Section 6.60)

Derivation of new time variables. The procedure for the derivation of the new time variables is set out in this section. The variables are based on the recorded periods spent by each child in the nursery and reception classes, during the period between the initial pre-test battery and the final post-test for that child. The diagram presented in the relevant section is reproduced again below.

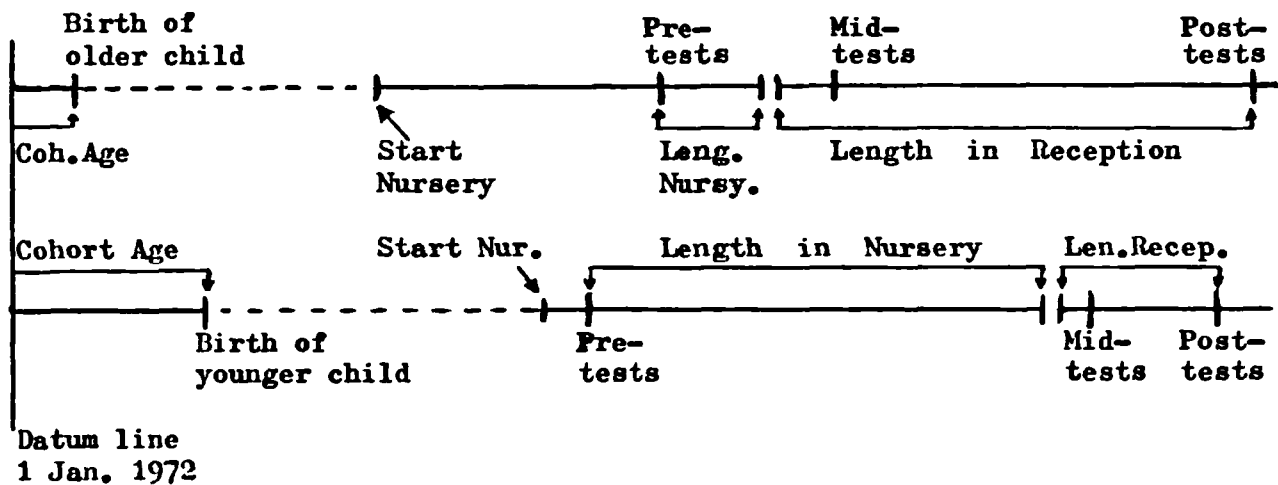


Figure 19. Diagrammatic representation of cohort age and time concepts

The 'length in nursery', 'length in reception' and cohort age variables are standardised for each school, and the standardised values are combined in equations yielding measures of 'time in nursery' and 'time in reception', freed from the age factor (which clearly affects the relative length of time spent by a child in the nursery or reception classes), as well as being adjusted for different school transfer practices. As a further precaution against misconstruing the importance of these variables, age variables representing age of testing are entered at each stage in the path model, including an entry of the raw variable 'age at post-test' in the final stage. In three of the ten models the variances in the 'time' variables prove to be stronger predictors than the variance of age at post-test. While a certain caution is still need in interpreting the value

of this new concept of Time in Nursery and Time in Reception, there is enough evidence in the path models to suggest that the variance in these variables does offer a quantifiable estimate of the contribution of both the nursery and reception curricula to academic performance. (Section 6.63)

Predictive power of meetings attended. The grave limitations of correlation coefficients on their own have been stressed at several points in this study. Apart from their real value as basic building blocks for multivariate models of relationships, they do have some limited value for adding to evidence from other sources. A stringent test was therefore carried out on some of the variables fundamental to an assessment of the effectiveness of the parent programmes, using both correlational evidence and direct regressions to test whether attendance at reading or mathematics meetings differentially influenced the post-test performance in reading and mathematics respectively. The test was particularly rigorous in that it was based on the whole sample, advantaged and disadvantaged, and included all those parents (over 35 per cent) who were either Working Group parents or who had failed to attend the meetings despite having accepted the invitations to do so. The evidence showed that attendance at reading meetings (by less than 40 per cent of the total sample), using two different criteria of attendance, correlated 0.11 and 0.10 with final reading performance for the whole sample; on the other hand attendance at mathematics meetings (by only 25 per cent of the sample) correlated slightly negatively with reading performance for the same sample. The reverse pattern existed for the prediction of post-test mathematics performance, with mathematics meeting attendance correlating 0.09 and 0.10 with maths performance and reading meeting attendance correlating only 0.03 and 0.04 with the maths outcome, again for the whole sample. Regression coefficients, from equations in which all four meeting variables were entered, confirmed the specificity of the effects of attending one or other meeting, even though these effects were seriously diluted by the nature of this test in which all the sample groups were combined. For the disadvantaged reading or maths groups on their own much higher correlations were found between meeting attendances and post-test outcomes, yielding 0.30 for reading meeting attendance and the Infant Reading Test, and 0.33 for maths meeting attendance and the Maths Concepts score; this compares with cross-correlations (reading attendance with maths outcome and maths attendance with reading outcome) of only 0.075 and 0.21 respectively. (Section 6.73)

Path..... /

Path Analyses. The principles underlying the path analyses in this study are set out here. The path models and path coefficients are based on a linked series of multiple regression analyses rather than on the interpretation of partial and simple correlations. The criteria for acceptance of a variable within a regression equation are a combination of the size or meaningfulness of the unique variance contribution of that variable, and the probability or significance of the regression coefficient for the variable. The use of V-ridge regression enables greater reliance to be placed on the regression parameters than would be the case with ordinary least squares regression, even with the moderately sized samples (from 27 upwards) on which the path models are based. Cross-validation evidence from many of the regressions confirmed the continuing superiority of the V-ridge solutions. Other criteria for the path model regressions include: simultaneous entry of all variables within a conceptual group or at a certain temporal stage (Nursery, Reception entry, or Post-Test), followed by a rejection of the poor predictors and a regression on the remaining predictors, until a satisfactory equation is obtained (thus automatic stepwise and hierarchical techniques are not used); the 'minor perturbation' of matrices which do not invert, in order to compel inversion in highly multicollinear situations where it is necessary to enter all the predictors initially, prior to rejection according to the formal criteria (after which perturbation is usually no longer needed); reliance on correlation rather than covariance matrices for the regressions; and the use of standardised data. The rationale for all these criteria is explained. (Section 6.80 et seq.)

'Suppressor variables'. The claimed phenomenon of 'suppressor variables' is examined with the aid of the regressions carried out in the study. This phenomenon is said to occur when a variable yields a regression coefficient opposite in sign to that of its correlation. A comparison of V-ridge and ordinary least squares (OLS) regression results shows that the phenomenon is simply another aspects of the OLS regression response to a multicollinear situation. In these situations the V-ridge regression coefficients are small or near zero and usually (though not always) have the same sign as the correlation coefficients. This finding raises serious doubts about the whole theoretical edifice which various authors have erected in an effort to explain the phenomenon. In essence, it appears that the phenomenon is simply an erroneous regression response to a particular type of multi-collinearity, with capitalisation on error (particularly in the OLS solutions), raising the level of a prediction which should probably be close to zero because it is largely duplicating other predictors. (Section 6.817)

Detailed construction (of path models). The method of building up the path models, and the role of latent and status variables, is set out in detail. (Section 6.822)

Procedures for inclusion and exclusion. The procedures for inclusion and exclusion of variables are set out in detail. It is shown that V-ridge regression coefficients are far more consistent, when poor predictors are removed from a model, than are the equivalent OLS coefficients. In the V-ridge solutions the coefficients and probability levels do not change radically and the proportion of unique variance predicted by each of the remaining variables is generally preserved. The parameters of the parallel OLS solutions, in contrast, vary considerably and at times there is a change not only of an order of magnitude but also in the sign of some of the OLS regression parameters. V-ridge is also seen to share out more of the variance among the predictors, and does not allocate nearly all the unique variance prediction to the largest predictors, as OLS tends to do. (Section 6.824)

Satellite models. The use of satellite models is explained. As it is desirable to compare directly certain sub-samples, such as boys and girls, or Black and White ethnic groups, it was decided to base the satellite analyses on the path model for the total disadvantaged sample of 129. This enabled the latent and status variables to be similarly constructed (using the same regression weights) across all the satellite models. On the other hand the varying path (regression) coefficients for each model served to assess the ways in which the samples differed from each other. This technique has its strengths and limitations, as explained in detail. Five of the ten models are satellites based on the total disadvantaged model, while the remaining four models together with the total disadvantaged model itself give five independent models for examination and comparison. (Section 6.825)

Limits set for probability and unique variance. The formal limits set for the probabilities of the regression coefficients and the minimum required levels of unique variance contributions are specified, prior to the creation and assessment of the path models. (Section 6.832)

#### PATH MODEL 1

The main path diagram shows the prediction of the total attainment at post-test, this outcome being a composite of each child's scores on five different reading and mathematics attainment measures at the end of the study period. The sample group for this model consists of all 129 children in the five schools in disadvantaged areas (excluding one child whose parents refused to take part in the programmes). A further diagram (section 7.20, p.661) presents the path model predictions in the form of a bar chart. The principal findings from this model include the following:

- More than 70 per cent of the variance of final attainment is predicted in this model.
- Out of the seven parent interview variables, only reading behaviours and maths behaviours (that is, the reported parental activities which are hypothesised to foster reading and mathematical skills or interests in the child) predict to the child's final total attainment some 20 months later, at the end of the study's field work. While parent reading attitude - a measure of the parents' expressed belief in the importance of reading to the child - correlated 0.51 with parent reading behaviours - the attitudinal variable has a near zero correlation with final attainment. This finding supports the decision in the present study to concentrate on evidenced or reported behaviours rather than on reported attitudes, which may well be at variance with behaviours in sensitive areas of child-rearing.
- The two most important predictors of final attainment are the child's level of initial academic attainment, as measured by the nursery class tests of reading and mathematical awareness, and the child's cognitive ability.
- The parent reading and maths behaviours, in the absence of other predictors, contribute 9 per cent to outcome variance (i.e. the variance in total attainment at the end of the study). However the parent environment variable makes no independent nor significant contribution when entered into the path model in competition with the child's initial attainment and ability variables; this is an important but also a disturbing finding.
- Despite the relatively high correlation of the English Picture Vocabulary Test with final attainment (0.51), E.P.V.T. makes no independent nor significant prediction of the academic attainment outcome for the disadvantaged sample.
- Both Time in Nursery and Time in Reception (the creation of these variables has already been described earlier in this summary) make small but significant contributions to final attainment. While the value of nursery education has been called into question in public debate, and its limited contribution to later academic attainment has been analysed in other studies reported in the review of the literature, the present study does indicate that nursery education is making a useful though moderate independent contribution to final attainment for a fairly representative disadvantaged sample.
- Attendance at either reading or maths programme meetings does not make a significant contribution to total attainment in this model, since the sample includes both attenders and non-attenders; however separate models for the disadvantaged reading and maths programme groups (4 and 6) show a significant prediction of reading or mathematics attainment respectively.



## PATH MODEL 2

The main path diagram shows the prediction of total attainment for the 27 children at the advantaged school (excluding two children whose parents refused to participate in the programme). A further diagram (section 7.20, p.661) illustrates the relative strengths of the path model predictors and enables a direct comparison of predictive patterns between the advantaged and disadvantaged samples. The findings include the following:

- Some 74 per cent of the variance of final attainment is predicted here, nearly all of this being predictable from the variables measured at the start of the study.
- There are considerable differences in the predictors and the strengths of these predictors, between the disadvantaged sample (model 1) and this advantaged sample.
- In contrast to the disadvantaged model, the parent academic environment makes a considerable contribution to outcome - larger than the child's initial attainment or ability.
- The E.P.V.T. (language) scores contribute meaningfully to final attainment, in sharp contrast with the failure of E.P.V.T. to predict in the disadvantaged model. A fuller study of the differences in prediction of this vocabulary measure, across the two models, shows that E.P.V.T. in the advantaged model also predicts to the child's cognitive ability at mid-test, whereas this is not found in the disadvantaged model. These and other findings on the differences in the predictive power of E.P.V.T. (already identified earlier in the discussion of nomological validities) suggest that the disadvantaged child's general word comprehension skills (as measured by this test) are not yet closely linked to the specific word knowledge required for adequate performance in the cognitive and academic tasks assessed by the tests. On the other hand the advantaged child's general language repertoire is likely to reflect also its specific word knowledge, judging by the relatively high contribution of the advantaged home environment to ultimate academic attainment. Further study is needed of these findings.
- The Time in Nursery and Time in Reception variables do not make any contribution to this model, suggesting that for advantaged children the academic skills and awareness they bring to nursery and reception class are such that these early years do not add noticeably to the variance in their final academic attainment towards the end of reception class, although, as pointed out earlier, there are clearly a great many social and other skills which are being learned by the advantaged children during this early period.
- The high shared variance at each stage in the model suggests a far more inte-

grated development of advantaged children than is evident in the disadvantaged model. This suggests that the skills of the advantaged sample are being developed in closer harness than is the case with disadvantaged children. The implications of this finding also need further consideration.

- The attendance at the reading programme meetings makes a modest contribution to outcome variance for the advantaged sample, even though less than half the sample group attended reading meetings; on the other hand a subsidiary regression shows that attendance at programme meetings is highly predictable from the variable reflecting the parent academic environment, indicating that for advantaged parents the study programme merely increased the home environmental factors already favouring the child's academic progress.

### PATH MODEL 3

The main diagram shows the prediction of total attainment for the children of working group parents at the disadvantaged schools. This is a satellite model, so that the creation of the latent variables and larger 'status' variables at each stage in the model is based on the coefficients derived from model 1; naturally it is the working group scores which are used entirely in the model. The findings below are based on the path coefficients derived from the data on the 28 dyads in this group. These findings include the following:

- In many ways the characteristics of the group are mid-way between those of the advantaged and disadvantaged groups, suggesting that working group parents in the disadvantaged sample tend to be socially upward mobile; evidence cited elsewhere in the study supports this contention.
- The children's initial academic attainment has a much larger predictive power than their early cognitive ability has, whereas for both the disadvantaged and advantaged samples the two groups of variables have equal predictive power. This suggests a particular focus on early attainment in the working group, possibly linked to a greater 'achievement press'.
- The parent academic environment makes a modest contribution to outcome variance.
- Both Time in Nursery and Time in Reception contribute to outcome at a higher level than for the disadvantaged sample as a whole, again suggesting that in some way working group parents are capitalising on the opportunities offered to their children by the nursery and school environments

#### PATH MODEL 4

The main diagram shows the prediction of post-test reading attainment (a composite of the scores on three reading tests) for the 48 children whose parents took part in the reading programme given at the five disadvantaged schools. The relative strengths of the predictors are shown in a further diagram on page 662. The findings include the following:

- While the prediction of final reading is only 50 per cent at the initial stage of the study, this rises to 67 per cent by the end of the study, suggesting that for this disadvantaged group a considerable part of the variance in final reading resides in the children's school experiences, including motivation of the children, combined with the effects of the parent reading programme.
- The rhythmic tapping test, a measure of auditory integration, proves a more powerful predictor of final reading than do any of the other individual (raw) cognitive measures, including the intelligence scales.
- The parent academic environment makes no contribution to the reading outcome, suggesting that the home behaviours of disadvantaged parents in this sample are not strongly related to the child's learning of reading at school.
- Both Time in Nursery and Time in Reception make important though moderate contributions to final reading, confirming the findings on the total disadvantaged sample that the nursery and reception class experiences do help to predict to final academic attainment at schools in disadvantaged areas.
- Even in this disadvantaged model focused specifically on the prediction of reading attainment, E.P.V.T. makes no contribution to final outcome. This suggests that early word knowledge is not the overriding factor in developing initial reading skills, at least for disadvantaged children, but rather that it is the mixture of cognitive abilities, initial attainment, motivational factors and learning opportunities which contribute most to early reading. It is fairly certain however that as the basic reading skills are mastered so does the level of language comprehension, including word knowledge, become increasingly vital for a fuller exploitation of the early skills and the development of true reading competence.
- One of the most important findings in this model, and indeed in the study as a whole, is that the parent reading programme does make a small though meaningful and significant contribution to final reading. This occurs in competition with the other predictors, with the contribution remaining constant when parent programme is entered at two successive stages in the model. A regression of reading programme attendance on the set of variables measured at the start of the study, including parent academic environment, shows that attendance at the programme is not in any way related to the parent environment. This is in con-

trast with the finding with the advantaged group, where it was shown that attendance at meetings was highly predictable from the initial parent environment. The implication is that parents at the disadvantaged schools who attended the reading meetings, and whose children benefited from the parent participation in them, were not influenced by the extent to which the parents were already showing or failing to show home behaviours likely to foster reading (or mathematical) competence in their children. Although the percentage of unique variance contributed is only 1.7 and 1.6 at the two successive stages in the model, these figures are approximately one-third of the contribution of Time in Nursery (at the mid-test stage of the model) and Time in Reception (at the final stage), indicating the particular usefulness of the eight one-hour parent meetings attended by this group.

#### PATH MODELS 5 AND 6

The main diagrams show the prediction models for the 32 parents attending the maths programme at the disadvantaged schools. For model 5 the predicted outcome is the final (post-test) score on mathematics numeracy; for model 6 the outcome is the final score on mathematical concepts. The findings include the following:

- There are considerable differences between the two models, reflecting the rather different nature of the predicted outcomes. While the total variance explained rises from 43 per cent at the initial stage to 53 per cent at the final stage of the path model for maths numeracy, for maths concepts this variance increases from 61 to 68 per cent. The lower level of explained variance in the numeracy model means that there are more unexplored predictors of numeracy here than there are of maths concepts.
- Ability is seen to count for much more of the variance in maths numeracy than initial attainment does, suggesting a strong ability component in early numeracy skills. In contrast early attainment counts for a good deal more than ability does in predicting maths concepts.
- The parent academic environment makes virtually no contribution to final numeracy, but does make an important though modest contribution to the final maths concepts score. An examination of the construction of the parent environment variable shows that the only meaningful components of this environment (as constructed separately for the maths concepts model) are the parents' 'maths behaviours' - the extent to which number related activities occur in the home - and a measure of television viewing time, the latter predicting negatively so that the more the viewing the lower the concepts score.
- Neither Time in Nursery nor Time in Reception make any meaningful contribu-

tion to either of the maths models. This appears to confirm the subjective evidence that the main form of activity in most of the nurseries, and to a lesser extent in most of the reception classes, is focused on reading related and other activities not having a strong mathematical content.

- The most important finding in these models is that the parents' attendance at the meetings of the mathematics programme makes a small but meaningful and significant contribution to the children's post-test scores on maths concepts, but makes virtually no contribution to post-test numeracy. For maths concepts the variance predicted is 2.8 and 2.0 at the second and third stages of the path model. Again the evidence also shows that attendance at parent programme meetings for the disadvantaged parents is not in any way predicted by the level of mathematical behaviours in the parent or home environment.
- The diagram on page 662 compares the relative strengths of the path predictors of reading with those of mathematical concepts.

#### PATH MODELS 7 and 8

The main diagrams show the prediction models for the 71 girls and 58 boys at the disadvantaged schools, with total post-test attainment as the outcome variable. Again these are satellite models, based on the initial disadvantaged sample (model 1). A further diagram on page 663 compares the relative strengths of the predictor variables for girls and boys. The findings include the following:

- The boys' parent academic environment is a very small but meaningful predictor of final attainment, whereas the combination of parent reading and mathematical behaviours makes no contribution to final attainment for girls. This suggests that even in a disadvantaged home situation where there is very little emphasis on fostering academic skills, boys enjoy at least some parental academic encouragement, whereas most girls do not have this particular form of stimulation.
- A contrasting finding is that Time in Nursery and Time in Reception predict considerably more strongly (though still modestly) for girls, than they do for boys. It is possible to relate this to the major problems of behavioural control and socialisation which arise with many of the boys at nursery level, and to a lesser extent in the reception class, as a result of which the management of these behaviours occupies a good deal of the time in which teachers might otherwise have interacted more usefully with the boys. Girls, on the other hand, tend to be better behaved and respond better to teacher stimulation, taking a greater interest in the more academic classroom activities such as looking at picture books and listening to the teachers' reading; they also tend to concentrate longer on activities.

### PATH MODELS 9 AND 10

The main diagrams show the prediction models for 44 Black children and 62 White children at the disadvantaged schools, with total post-test attainment as the outcome. This too is a satellite model. A further diagram on page 664 compares the relative strengths of the path model predictors of Black and White children. The findings include the following:

- The variance in final attainment explained in the Black model rises from 52 per cent at the nursery level to 63 per cent at the post-test level, whereas the variance explained in the White model rises from 72 per cent at the outset to 76 per cent at the final stage. This considerable difference suggests that the nursery and reception experiences of the Black children may have been more productive than they were for the White children.
- The contributions of Time in Nursery and Time in Reception appear to confirm the above finding, namely that these variables make a larger prediction for Black children than they do for White children, although for both ethnic groups the contributions are satisfyingly large, pointing to the value of the nursery and early school experiences for these groups at schools in disadvantaged areas.
- Although the parent academic environment makes no independent prediction for either the Black or White groups, in accordance with similar findings for the total disadvantaged sample, the regressions of the outcome variable on to the individual parent home behaviours, in the absence of other predictors of attainment, show that White parents' academically-related behaviours make a more meaningful contribution to final attainment than do the Black parents' behaviours. This confirms the earlier findings on the differences in the mean levels of the parent behaviours across the ethnic groups. On the other hand the models also show, over the three stages of each model, that the Black children's cognitive abilities contribute considerably more to final attainment than do the White children's abilities. For the White children it is the initial attainment levels which make a higher contribution to final attainment than they do for Black children, again suggesting that Black children reach nursery with fewer of the pre-academic skills required by the school, but having equal levels of cognitive skills with those of the White children and relying to a greater extent on these skills for early academic attainment.
- For Black children the Needs variables make a very small but meaningful contribution to outcome, whereas this is not the case in the White model. The implication of this finding could be quite serious, as it suggests that Black children entering a White dominated environment for the first time may feel a lack of self-esteem and insecurity in what is for them a doubly strange environment - school rather than home, and White rather than Black. This is not to suggest any lack of concern or compassion on the part of the teachers and nursery

nurses - whose numbers included Asian staff and a Black deputy head - but it may reflect the subjective perceptions by Black children of the school situation.

• The above discussion of ethnic differences in the home academic environment needs to be interpreted against a background of the traumatic past history of the Blacks, particularly of the West Indians who form the majority in this sample; their historical experiences may well have influenced the cultural self-concepts, skills and customs which underlie the degree to which Black parents are willing to attempt to influence their children's development and to prepare them for schooling in a culturally alien society. It is important to note moreover that the ethnic differences in the 'academic' environment of the homes exist alongside an equality in the competence, at initial nursery level, of the children in the two ethnic groups.

• These findings may have a bearing on the current debate over the causes of poor school performance in West Indian children, compared to White and Asian children. One major British study has suggested that the schools should endeavour to change Black parents' attitudes towards and trust in the schools, and that staff should also place greater emphasis on building self-esteem among Black children. Such viewpoints ignore the fact that a parent's education-related attitudes and a child's self-esteem can only arise from a foundation in which the home's educational 'behaviours' and experiences have already provided the pre-school child with the opportunity for success in early educational tasks, within a situation of joint parent-child interaction. The critics' exclusive focus on the need for institutional change neglects the greater possibilities of persuading parents to create a strongly educational environment in the home, with actual behaviours rather than attitudes as the prime target of parent guidance.

Cost-benefit indicators. Estimates have been made, subject to a number of assumptions, of the costs and benefits of three contrasting programmes:

- a. the two separate study programmes, as these have been carried out here (with a fairly high cost in time of preparing the programme materials);
- b. a proposed regular nursery programme, which could be run for small parent groups at all nursery classes within disadvantaged areas, providing two-hour monthly meetings throughout the nursery 'year' and offering guidance on both reading and mathematical games and activities, with programme materials produced centrally to reduce costs; and
- c. a typical remedial reading programme, given for poor readers at junior school level, making a generous assumption that this programme would result

in an average improvement of one standard deviation in reading scores on a standardised test.

The estimated expenditure in teacher time and the estimated results of these programmes are set out below.

Table 42. Comparisons of the cost-effectiveness of three intervention programmes

	Study programmes		Regular Nursery prog.		Remedial Reading in Junior Sch. (6 per grp.)
	Reading (3-6 parents per gp.)	Maths (3-6 parents per gp.)	Combined Reading/Maths (3-6 parents per grp.)		
Number of parents	48	32	80		80
Frequency meetings	Fortnightly 1 hour	Fortnightly 1 hour	Monthly 2 hours		2 hours per week
Duration	8 meetings	8 meetings	7 meetings		6 terms (2 yr)
Programme meetings and resources (teacher hours)	198	132	<div>264</div> <div>132 (reading)   132 (maths)</div>		2240
Effect of 1 standard deviation increase in programme length	0.20 std. dev. reading improvt.	0.29 std. dev. maths improvt.	0.20 std. dev. reading improvt.	0.29 std. dev. maths improvt.	0.4 std. dev. reading improvt.
Estimated cost of incr. of 1 std. dev. in programme length (teacher hours)	57.2	38.1	38.1	38.1	258.7
Estd. cost of incr. in programme length per parent-child dyad (teacher hrs.)	1.19	1.19	0.48	0.48	3.23
Estd. cost per dyad of 1 std. dev. in improvt. in reading or maths concepts (teacher hours)	5.95	4.10	2.40	1.67	8.08

Subject to the various assumptions described in the relevant section, these comparisons indicate that even given a generous estimate of the effectiveness of the remedial reading programme, parent programmes are likely to be more cost-effective than later remedial programmes focused on the child. The advantage of the parent programmes may be greater still if the long-term effectiveness of these programmes is taken into account. (Section 6.90)



## 7.20 Comparison of variance predictions

On the following pages the contributions of different variables to the children's final attainment scores are set out in the form of bar charts. The unique variance contributions have been taken from the path models derived in chapter 6. In view of the considerable differences in shared variance across the different models the procedure has been adopted of dividing out the shared variance among the predictors, in exact proportion to the unique variance contributions of each predictor. Thus, for example:

### Prediction with low shared variance. (Second level of path model 1)

Total variance predicted at this level            71.2 %  
Shared variance at this level                        4.8 %

Predictor Variable	Unique Var. Contrib.(%)	Fraction of Comb. Uniq. Variances	Proportn. of Shared Var. attributed	Total: Unique + Attributed Variance (%)
Initial Attainment	11.1	0.167	0.80%	11.9
Time in Nursery	3.6	0.054	0.26	3.9
Nursery Ability	1.5	0.015	0.07	1.6
Reception Ability	12.3	0.185	0.90	13.2
Nursery Status	38.0	0.571	2.74	40.6
Totals	66.5	c. 1.0	c. 4.8	71.2

### Prediction with high shared variance. (Third level of path model 1)

Total variance predicted at this level            71.2 %  
Shared variance at this level                        34.7 %

Predictor Variable	Unique Var. Contrib.(%)	Fraction of Comb. Uniq. Variances	Proportn. of Shared Var. attributed	Total: Unique + Attributed Variance (%)
Initial Attainment	7.2	0.197	6.84	14.0
Time in Nursery	0.4	0.011	0.38	0.8
Nursery Ability	2.0	0.055	1.91	3.9
Reception Ability	7.0	0.192	6.66	13.7
Nursery Status	9.1	0.249	8.64	17.7
Time in Reception	1.5	0.041	1.42	2.9
Age at Post-Test	0.9	0.025	0.87	1.8
Reception Status	8.4	0.230	7.98	16.4
Totals	36.5	1.0	34.7	71.2

This compensating technique, as set out on the previous page, appears to overcome the problem of comparing individual variance contributions across levels or models in situations where the amount of shared variance alters radically between two multiple prediction equations in which many of the same variables occur. While there may be alternative methods of apportioning shared variance, the proportional division procedure appears simple and relatively uncontentious; shared variance is in reality made up of the common contributions of all the predictors, with the shared commonality of any one predictor likely to bear some reasonable relationship to the size of the unique variance of that contributor. The only predictors which would be non-contributors to shared variance would be those having a zero or near zero correlation with most if not all the other predictors; in none of the models does that orthogonal situation arise.

The comparisons of variable predictions set out in the bar charts are as follows:

Total Disadvantaged Sample      vs.      Advantaged Sample,  
predicting Total Attainment      p. 661

Disadvantaged Reading Groups      vs.      Disadvantaged Maths Groups  
predicting      predicting  
Reading Attainment      Maths Concepts Attainment      p. 662

Disadvantaged Girls      vs.      Disadvantaged Boys  
predicting Total Attainment      p. 663

Disadvantaged Black Children      vs.      Disadvantaged White Children  
predicting Total Attainment      p. 664

It should be noted that in a number of the models there are minor statistical anomalies due to a rigorous correction procedure according to which the more variables that are entered the more stringent is the enforced reduction in the total variance prediction. This correction procedure, which is aimed at avoiding the errors caused by chance inflation in the variance prediction, also reduces the size of individual variable predictions. For the sake of clarity in the bar charts, the anomalous effects of differential reductions in the predictive power of some variables at successive levels in a model have been smoothed to the level of the lowest predictions.

Figure 22. Comparison of variance predictions across path models: Disadvantaged Sample vs. Advantaged Sample (predicting Post-Test Total Attainment)

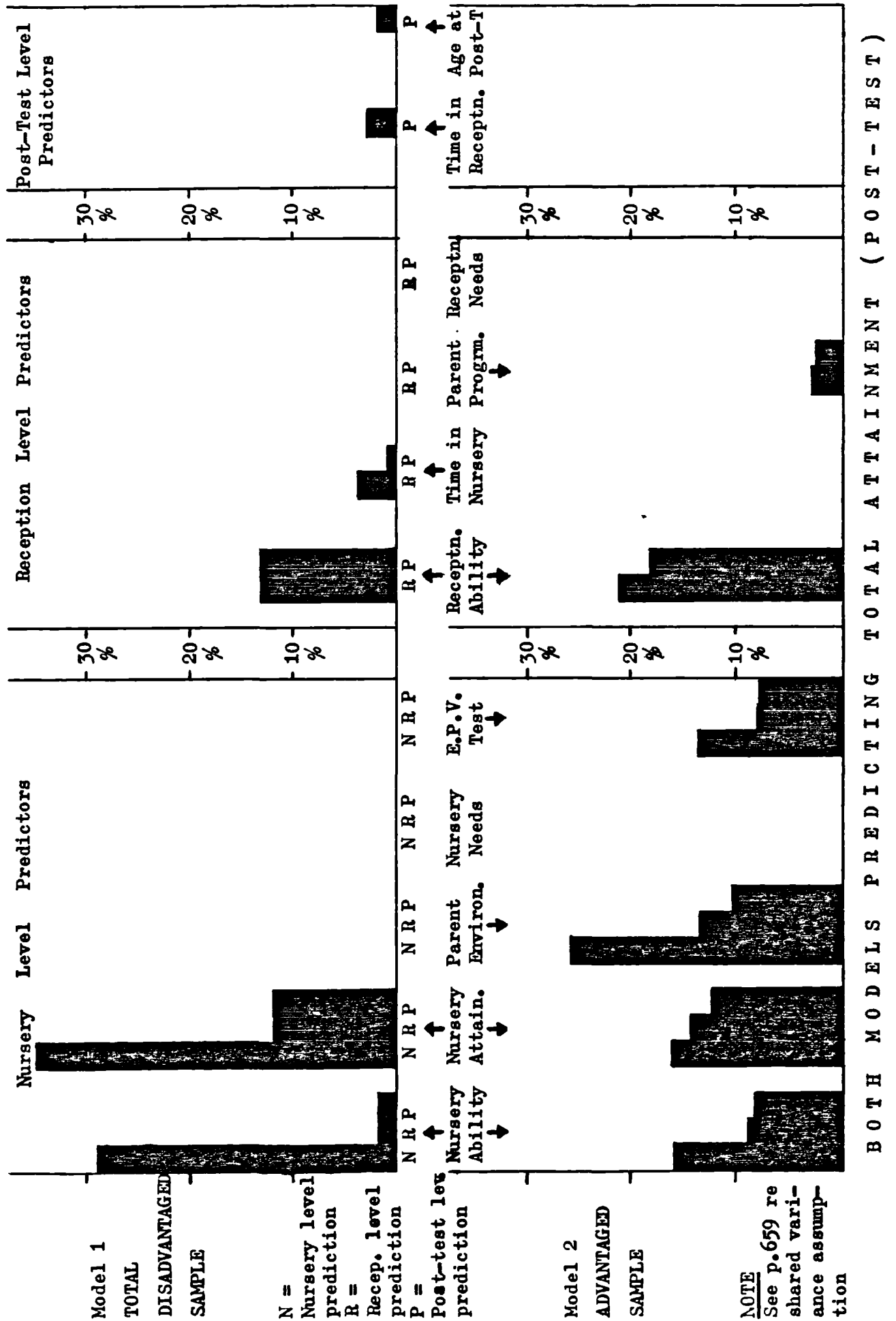


Figure 23. Comparison of variance predictions across path models: Disadvantaged Reading Groups (predicting Post-Test Reading) vs. Disadvantaged Maths Groups (predicting Post-Test Mathematics Concepts)

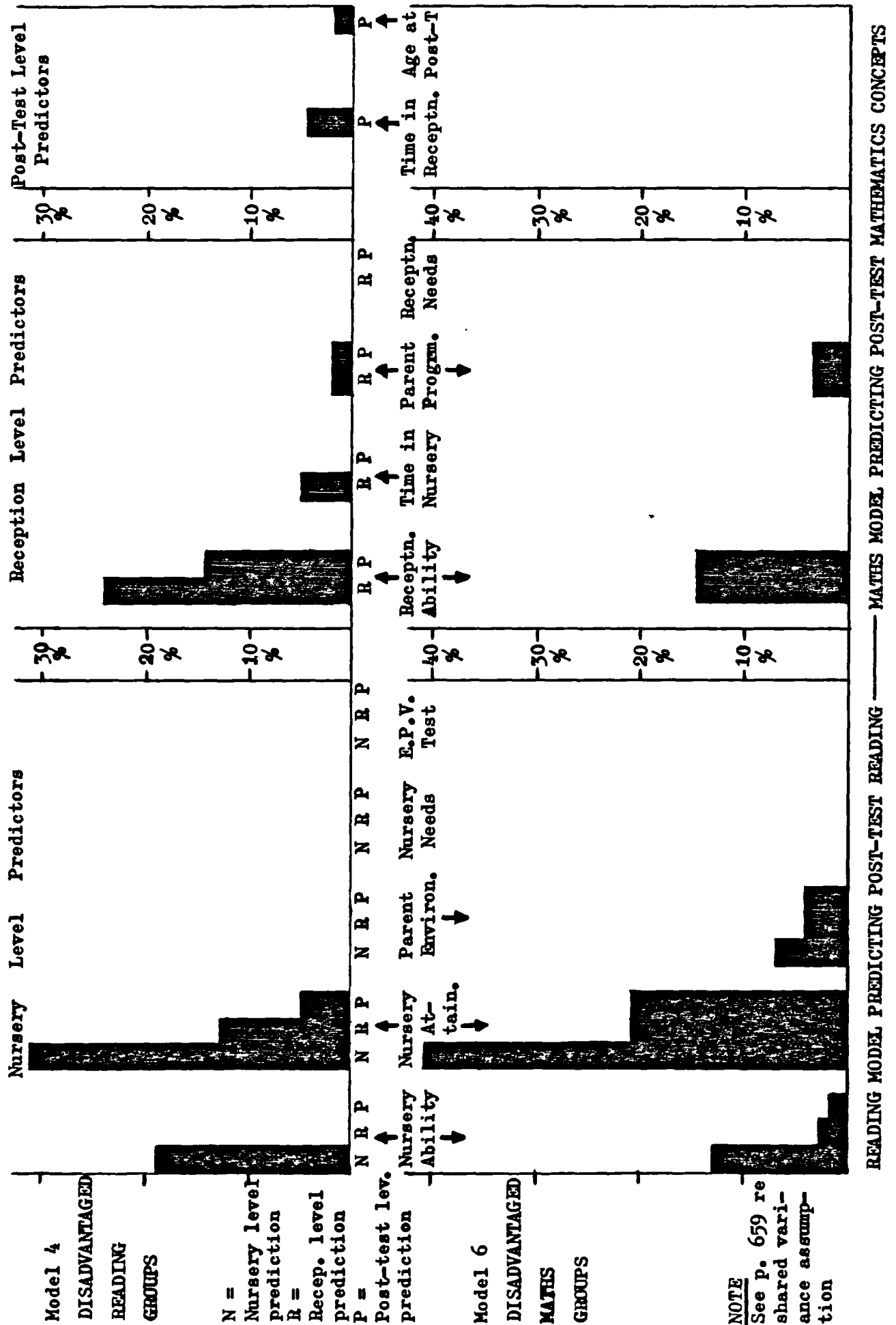


Figure 24. Comparison of variance prediction across path models: Disadvantaged Girls vs. Disadvantaged Boys (predicting Post-Test Total Attainment)

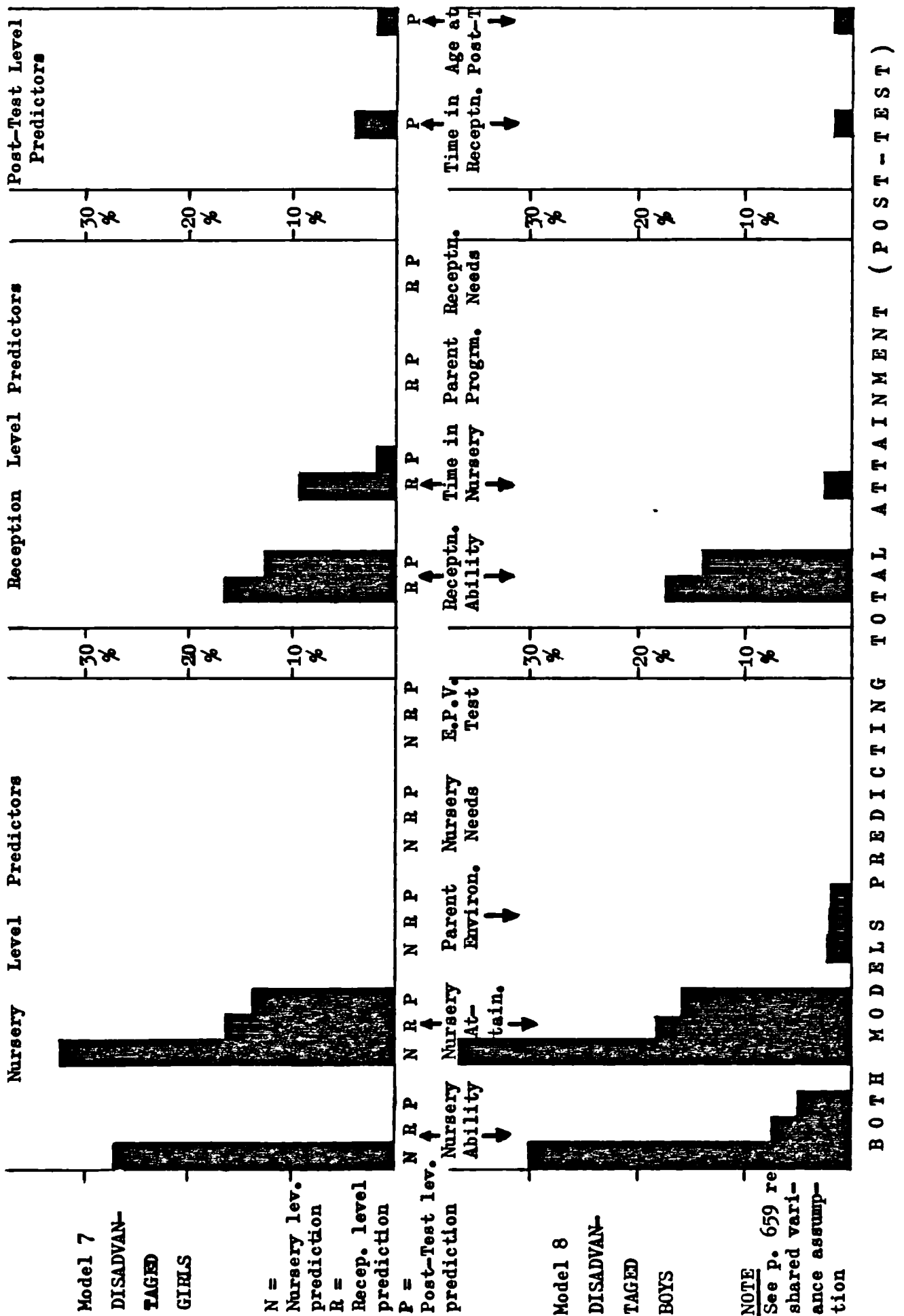
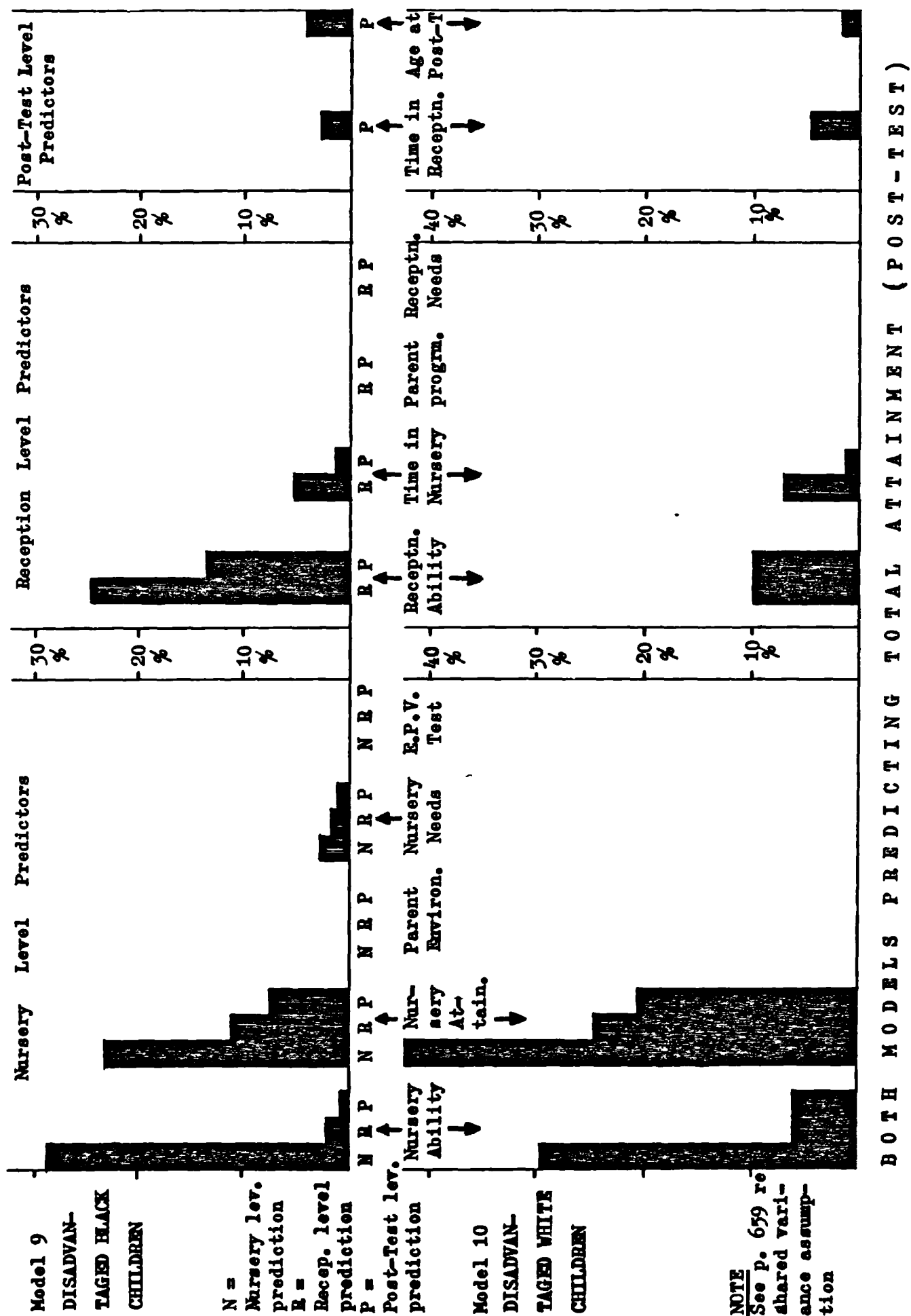


Figure 25. Comparison of variance predictions across path models: Disadvantaged Black children vs. Disadvantaged White children (predicting Post-Test Total Attainment)



### 7.30 Principal Findings

### The Study Hypotheses

- A. That the intervention programmes, focused on the parents of nursery class children, will differentially influence reading or mathematics attainment in the children's first year of school according to the nature of the programme given to different groups of parents.

**Finding:** Confirmed  
(Section 6.73)

- B. The parent participation in the intervention programme will prove to be a significant contributor to first year school attainment, beyond the contribution of the assessed 'parent academic environment', the child's pre-test academic skills and the child's pre-test cognitive skills.

**Finding:**            **Disadvantaged Sample:** Confirmed for reading programme;  
partially confirmed for mathematics  
programme, contributing to concepts  
but not to numeracy.

**Advantaged Sample:** Not confirmed

(Section 6.84: Models 2, 4, 5, 6 )

- C. That path models of the contributors to early academic attainment will show that there are important differences between the sexes, between ethnic groups and between social levels, in the particular variables which contribute to performance.

**Finding:** Confirmed  
(Section 6.84: Models 1, 2, 7, 8, 9, 10 )

### Findings of Potential Importance

**From the review of the research literature, chapters 2 and 3.**

1. Nursery schooling: the social bias of the curriculum? Although pre-schooling was started as an educational measure for the disadvantaged, it has gradually been adapted to the needs of the middle class and the orientation of the curriculum has changed from a concern for education, welfare and health to

a concern for social and emotional adjustment and creative expression.

2. Nursery schooling: supply and demand. Nursery provision in the United Kingdom continues to face rapid swings between expansion and contraction; meanwhile parents themselves are reported to find a mismatch between the values of the nursery school and the working world which their children must one day enter. The issues of both supply and demand are closely related to the uncertainty as to what are or should be the goals of pre-schooling.

3. Nursery schooling: research findings. While the social inequalities inherent in the free but laissez faire economies of the West mean that there are cogent reasons for providing pre-schooling, including special intervention programmes, in areas of disadvantage - assuming a close relationship between educational and social disadvantage - research in the United States and United Kingdom suggests that normal nursery schooling and also many intervention programmes have only limited results. Programmes which are well structured rather than being informal, and with a strong educational rather than creative play content, appear to achieve greater success in terms of the children's ultimate academic attainment.

4. Intervention: institutional or parent? A great deal of research suggests that pre-school parent intervention is considerably more effective and long-lasting than institutional intervention focused only on the children. Although many professionals are reluctant to enter the home, preferring to deal only with the children or to meet the parents in the security of the institution, the evidence points to the need to start intervention in the home, using sensitive home visitors, especially with those parents who normally fail to attend institutional programmes.

5. Intervention: research shortcomings. Among the shortcomings often identified in intervention research are: an over-emphasis on laboratory-type designs rather than field-oriented action research designs; insufficiently comprehensive data collections, and reliance on simple analyses rather than multivariate models which take more account of reality; the inevitable problems of adequate sampling; using the criterion of an increase in intelligence quotients rather than academic improvement as evidence of the success of what are essentially educational programmes; and an over-emphasis on assessment of the child in isolation, with a corresponding failure to assess adequately the child's total environment - the latter being as important a predictor of later attainment as are the child's own characteristics at an early age.

From the study...../



From the study programmes, their results and the analyses, chapters 4, 5 and 6.

6. Sample parents' views on the nursery curriculum and pre-school preparation.

The majority of nursery parents at the five disadvantaged schools in the sample considered that the nursery curriculum was aimed mainly at the socialisation of their children; other goals were thought to be general development and the fostering of language; only a minority felt that what was frequently termed 'play-school' had any educational content. Most parents did not want to teach their pre-school children any reading skills, in case they did it 'incorrectly' and 'damaged' their children. Very few parents made any use of public libraries.

7. Attendance at parent programmes. More than 200 children and their parents were assessed initially; 159 dyads were traced for assessment at the end of the field study. In 80 per cent of the sample at least one parent was not working outside the home, and 80 per cent of this 'non-working' group attended one or more meetings of either the reading or mathematics parent programmes given at the six sample schools. More than half the programme parents continued attending up to the end. The less confident and less competent parents were more likely to stop attending. Fears expressed by some teachers about undue parental pressure were not confirmed.

8. Parents' views on the programmes. A variety of issues were raised and proposals made for future programmes of this kind. For example, parents needed reassurance that children who benefited from such programmes would not suffer when they started in reception class. Parents appreciated the value of the programmes, and also the opportunity to interact with other parents, but they wished the programmes could be more adapted to individual needs, although the difficulty of doing this in group situations was recognised. The clear structuring of the programmes and the demonstrations were welcomed. However many parents would prefer monthly rather than fortnightly meetings.

9. Development of new form of regression analysis. A considerable amount of time was given to developing a new form of non-stochastic ridge regression for the multivariate analyses; this technique is superior to the ordinary method of regression, yielding more reliable and consistent estimators. Unlike normal ridge regression, it offers unique and repeatable results. The development of this technique, based on an algorithm pioneered by the American statistician Vinod, is regarded as a major contribution to future research analysis when dealing with 'soft' social data.

10. Differences between sub-samples. Comparisons of the scores of parents and children on a variety of measures showed considerable differences in the expected directions. The home environments of advantaged parents and also the cognitive

and pre-school academic attainment levels of their children were well ahead of the means for the disadvantaged dyads, with the advantaged children up to 40 per cent ahead on some variables. Girls' scores were superior to those of boys on most tests. There were virtually no differences between Black and Non-Black children on a number of key pre-test variables; on one test (copying complex rhythmic patterns) the Black children were far superior to Non-Blacks. On the other hand the parent interview scores on a number of academically related home behaviours showed White and Asian parents ahead of the Black parents on these measures; the attendance levels of Black parents at the programme meetings was also somewhat lower than those of other parents. However on every variable there were some Blacks at or near the top of the distribution, and some Whites at or near the bottom. The social and ethnic differences in the academic orientation of the home environments are closely related to cultural and historical factors.

11. Differences between parent programme groups. Comparisons of parent and child scores across the parent programme groups in the disadvantaged schools showed that working group parents had higher mean scores than parents who did not go out to work; these differences were also reflected in the children's scores. The reverse applied in the advantaged school, where working group scores were lower. The scores of programme attenders were above those of non-attenders. Although those who accepted the invitation to attend programme meetings were randomly divided into reading and maths groups, the scores of the reading and maths programme attenders showed a surprising bias in favour of maths attenders, reflecting the somewhat less popular nature of the mathematics programme.

12. English Picture Vocabulary Test. Despite the high correlations of E.P.V.T. scores with post-test attainment scores in reading and mathematics, this language (word knowledge) variable does not make any independent contribution to variance in either the main or subsidiary disadvantaged samples, suggesting that the measures of cognitive and academic attainment entered in the same models are already duplicating any information given by E.P.V.T. In contrast, E.P.V.T. makes a considerable independent contribution to post-test outcome scores in the small advantaged sample, alongside the other predictors. This puzzling finding has been examined in depth; it is suggested that while general word knowledge is crucial for the development of ultimate reading and mathematical competence, it is not as crucial for the initial stages of reading or mathematical development as are early attainment and early intelligence.

13. Time in Nursery and Time in Reception. Two variables created to reflect the relative time spent by sample children in the nursery and reception classes respectively, taking account also of the children's age patterns and individual school transfer practices, showed that both the nursery and reception experiences

made meaningful though modest contributions to post-test academic attainment for children at the disadvantaged schools. For advantaged children, however, these time variables did not contribute to post-test (academic) attainment, suggesting that such children reached school having already undergone so much academic formation in the home that the value of the early school years could be found more in the spheres of social and emotional development than in that of academic development.

14. Path analysis of Disadvantaged sample. The path analysis of the predictors of post-test attainment, based on pre-test and mid-test measures, showed that while the disadvantaged children's early cognitive skills and pre-test academic attainment both contributed handsomely to final academic attainment in a combined measure of reading and mathematics, the parent academic environment (as measured in the parent interviews) made no independent contribution to variance. The English Picture Vocabulary Test likewise made no independent contribution. However both time in nursery and time in reception made useful contributions.

15. Path analysis of Advantaged sample. This path model differed widely from that of the disadvantaged sample. While the advantaged child's cognitive and early academic attainments made useful contributions to final outcome, as they did in the disadvantaged model, the parent academic environment and the E.P.V.T. scores also made further important and significant contributions, in sharp contrast to the disadvantaged model. On the other hand time in nursery and time in reception did not make independent contributions. Another finding of importance was that the apparently useful contribution to outcome variance from the parents' attendance at reading programme meetings was strongly predictable from the parents' home environment score, suggesting thus only a limited value for the programme for this sample.

16. Path analysis of Disadvantaged Working Group sample. The path model suggested that the characteristics of this sample were midway between the disadvantaged and advantaged samples, pointing to a possibly upward mobile group. An additional finding was that post-test attainment for this group was more heavily dependent on the children's initial attainment and on time in nursery and time in reception than was the case for either the disadvantaged or advantaged samples.

17. Path analysis of Disadvantaged Reading programme groups. This model showed that attendance by the parents at the reading programme group meetings made a significant though modest contribution to final reading attainment. It is important to note that this attendance was not in any way predictable from the parents' home environment scores (although it was predictable in the case of the advantaged sample). Other evidence from this model suggested a considerable

additional contribution to academic attainment, over the 20 months of the study, from the nursery and reception experiences.

18. Path analyses of Disadvantaged Mathematics programme groups. These two models (based on the outcomes of mathematical concepts and mathematical numeracy respectively) showed that the parents' home environment made a significant though modest contribution to the variance of post-test mathematical concepts, but not to that of post-test mathematical numeracy. Parent attendance at the maths programme meetings also predicted meaningfully to post-test maths concepts, but again not to post-test numeracy, emphasising the importance of establishing first the foundation of concepts. Neither time in nursery nor time in reception made any independent contribution to post-test scores in maths concepts or numeracy, suggesting that the main focus of these nursery and early school experiences was on reading rather than mathematical development.

19. Path analyses of Disadvantaged Girls and Boys sub-samples. These models show two important differences between the sub-samples. The parent academic environment made a small contribution to final academic attainment for boys but not for girls, suggesting that parents in disadvantaged areas tend to offer some minimal academic type stimulation to boys but not to girls. On the other hand time in nursery and time in reception were more predictive for girls than for boys. This in turn suggests that girls make more constructive academic use of their early school experiences; in contrast, a good deal of teachers' time needs to be focused on the difficult task of establishing behavioural control among boys in those early years, at the expense of time that might otherwise be spent on more productive experiences.

20. Path analyses of Disadvantaged Black and White sub-samples. These models show interesting differences between the predictive strengths of the variables. Over the period of the study there is a much greater increase in the prediction of Black attainment than there is of White attainment. Both time in nursery and time in reception are of a higher predictive value for Black children than for Whites. Although in both the Black and White disadvantaged samples the parents' home environment is not a meaningful predictor in competition with the other variables, there is a higher correlation between final academic outcome and the parent home behaviours for Whites. The models also show that Black children rely more on initial cognitive skills for their final attainment, while White children rely more on their initial attainment levels. The levels of Black children's assessed needs for esteem and security, in what is for them a doubly alien environment, are also meaningful contributors to ultimate academic attainment, whereas this is not the case with White children.

21. Cost-benefit indicators. A simplified cost-benefit evaluation indicates

that parent programmes are more cost-effective than even an assumed highly successful remedial reading programme for junior school children. If the study programme's heavy costs in time of preparing the programme materials were to be shared between all the nursery classes within an educational authority, and with other indicated improvements in the efficiency of providing a combined reading and mathematics programme to small groups of nursery parents at seven monthly meetings during the nursery year, the cost-effectiveness of parent programmes would be up to four times higher than that of an equivalent remedial reading programme. This reinforces other findings suggesting that parents are the most economic teaching resource available to schools in a child's early years.

7.40      The future?..... /

7.40      The future ?

The question arises of whether the lessons learned and the evidence gained from the parent intervention programmes have any relevance for the future.

The evidence does indicate that parent programmes can be effective, even if the contribution to the children's early educational attainment appears limited in relation to the child's own characteristics and, potentially, to the home environment itself. All the path models indicate that the school experiences in the nursery and reception years are crucial for disadvantaged children, in terms of growing educational competence. There is every reason why parents should be recruited to help in this process.

Fears of 'undue pressure by parents', in a fairly large sample studied as fully as this one was, have been shown to be unfounded. On the contrary, the majority of parents are eager to learn how they might best help prepare their children for early schooling, though mistakenly fearful and unwilling to make the attempt in case they may not be following what they believe exists as the one 'right' method.

The findings from this study suggest that a parent programme run by each nursery class or by teachers from the adjoining infant school classes could provide an overlapping series of monthly meetings for small groups, with up to six in each group, ensuring that all nursery parents who are not already working outside the home receive guidance on developing early reading and mathematical skills in their children. In terms of cost-effectiveness such programmes are likely to be much less expensive in teacher time than remedial reading or remedial maths programmes when the children reach junior school and, in a disturbingly large minority of cases, become identified as being in need of remediation.

The evidence from the study also suggests that programmes such as these should be funded and focused mainly on disadvantaged areas, where the need for parent guidance and encouragement is at its greatest.

In the event that educational authorities (or the teachers themselves) do not see such programmes as basic to infant education, as customarily understood for children from 4 to 7, the alternative of recruiting para-professional aides could be considered. These aides could work in collaboration with the school authorities, but be funded so that they can organise the parent programme groups wherever these may be most appropriately arranged and housed. Even in the event that teachers accept this responsibility, the problem will still remain of those parents who are never reached by any kind of group programme, who would need to be visited in their homes in order to be offered help and ideas on stimulating

their children; for such parents the health visitor or para-professional aide, or the trained home teacher, may well prove to be ideal workers.

For both parents and children a pre-school educational programme such as that carried out for this study could be of great value. For the parents it can help to build self-confidence in their parenting skills, particularly at the time when their children are at the difficult interface between home and school and when an enthusiastic and academically aware parent can achieve as much or more than even the most skilled infant teacher. For the children it could be equally valuable, discovering that their parents share to some degree in the early educational environment which the children encounter at school, and being made aware that the home cultural environment - which is often unconcerned about educational matters in the face of poverty and other hardship - does nevertheless regard early attainment in basic reading and mathematical skills as the most important step in the child's development during the first school years.

When the professional teacher, or indeed any other professional, learns that the sharing of basic skills with parents is not to diminish the profound importance of the professional's educational role but rather to widen it and thereby draw increased respect for her higher skills, society may be on its way to moving out of a situation in which an increasing minority start as unprepared hopefuls and end as educational failures, often becoming a lifelong burden on the same society which ignored their early needs. What should be done is neither vastly expensive nor radically different from what common sense might suggest.

## References

- American Dev(1)elopment Mathl ematical Soc ial H.M.S.O. Her Maj.  
 Association Edu(1)cation Psy(1)chology Socy Society Stationery Offc.  
 British Int ernational Rdg Reading Sociol(1)ogy DES/SSHC Dept Edu  
 Bulletin J ournal Rev iew Stat istical & Science / Soc  
 Child Lang uage Rsh Research UP Univ.Press Sci Rsh Council
- Abt Associates (1977) Education as experimentation: an evaluation of Follow Through Abt Associates Cambridge Mass.
- Achenbach T M and Weisz J R (1975) Impulsivity-reflectivity and cognitive development in preschoolers: a longitudinal analysis of developmental and trait variance Devl Psy 1 413-414
- Acland H (1971) What is a 'bad' school? New Society 9 September 1971
- Ainsworth M, Bell S and Stayton D (1974) Infant-mother attachment and social competence in Richards M (Ed) The growth of competence Academic
- Ambron S R (1977) A comparison of infant and parent education programs in Day M C and Parker R K (Eds) The preschool in action Allyn and Bacon
- Ammon P (1977) Cognitive development and early childhood education: Piagetian and neo-Piagetian theories in Hom H L and Robinson P A (Eds) Psychological Processes in Education Academic
- Anastasi A (1976) Psychological Testing Macmillan
- Anderson J A (1976) Multivariate methods in clinical assessment J Royal Stat Socy Series A 139 161-182
- Anderson R B et al (1978) Pardon us, but what was the question again? A response to the critique of the Follow Through evaluation Harvard Edul Rev 48 2 161-192
- Anderson S (1973) Educational compensation and evaluation: a critique in Stanley J C (Ed) Compensatory education for children ages two to eight John Hopkins U P
- Anderson L and Settle R (1977) Benefit-cost analysis: a practical guide Lexington
- Andison F S (1978) Public Opinion Quarterly 41 3 p.314
- Anscombe F J and Tukey J W (1963) The examination and analysis of residuals Technometrics 5 141-160
- Applebee A (1971) Research in reading retardation: two critical problems J C Psy and Psychiat 12 91-113
- Armstrong G and Brown F (1979) Five years on Social Evaluation Unit Oxford
- Athey I (1970) Affective factors in reading in Singer H and Ruddell R B (Eds) Theoretical models and processes in reading Int Rdg Ass
- Atkinson D (1976) Learning from others: two contrasting patterns of nursery education Edu 3 to 13 4 1 53-55
- Ault R L et al (1976) Some methodological problems in reflectivity-impulsivity research C Dev 47 227-231
- Austin G R (1976) Early childhood education - an international perspective Academic
- Badia P, Haber A and Runyon R P (1970) (Eds) The null hypothesis: acceptance/rejection strategies Addison-Wesley
- Bakan D (1970) The test of significance in psychological research in Morrison D H and Henkel R E (Eds) The significance controversy Butterworth
- Bakker D J (1972) Temporal order in disturbed reading Rotterdam U P
- Baltes P B (1968) Longitudinal and cross-sectional sequences in the study of age and generation effects Human Dev 11 145-172
- Banks O (1968) The sociology of education Batsford
- Banta T (1970) Tests for the evaluation of early childhood education in Hellmuth J (Ed) Cognitive studies Brunner/Mazel
- Barker R (1965) Explorations in ecological psychology Am Psychologist 20 1-14
- Barker R and Wright H (1954) Midwest and its children Row, Peterson Evanston



- Barker W E (1976) The development of a motivational test battery and its relationship to early reading M A Dissertation Institute of Education U London
- Barker W E (1983) Intervention and child development: health, psychological and other concepts Colston Symposium U Bristol March 1982
- Bartko J (1966) The intraclass correlation coefficient as a measure of reliability Psyl Reports 19 3-11 (and corrective note Psyl Reports 1974 34 418)
- Bartko J (1976) On various intraclass correlation reliability coefficients Psyl Bulletin 83 762-765
- Bartko J and Carpenter W (1976) On the methods and theory of reliability J Nervous Mental Disease 163 307-317
- Baumrind D and Black A (1967) Socialisation processes associated with dimensions of competence in preschool boys and girls C Dev 38 291
- Bayley N (1955) On the growth of intelligence Am Psychologist 10 805-818
- Beaton A E et al (1976) The acceptability of regression solutions: another look at computational accuracy J Am Stat Ass 71 158-171
- Becker G (1975) Human capital Columbia U P
- Becker G and Chiswick B (1975) Education and the distribution of earnings: a statistical formulation in Becker G (Ed) Human capital Columbia U P
- Becker W (1977) Teaching reading and language to the disadvantaged: what we have learned from field research Harvard Edul Rev 47 518-543
- Becker W and Carnine D (1978) Direct instruction - a behaviour theory model for comprehensive educational intervention with the disadvantaged Eighth Symposium on Behaviour Modification, Venezuela
- Bell L (1969) An experiment in home visiting practice Mimeographed DES/SSRC
- Bennett N (1976) Teaching styles and pupil progress Open Books
- Bercovici A and Feshbach N (1973) Teaching styles of mothers of 'successful' readers and 'problem' readers in the first grade Am Edul Rsh Ass New Orleans
- Bereiter C (1972) An academic preschool for disadvantaged children: conclusions from evaluation studies in Stanley J C (Ed) Preschool programs for the disadvantaged John Hopkins U P
- Bereiter C and Engelman S (1966) Teaching disadvantaged children in the preschool Prentice-Hall
- Berg I (1970) Education and jobs - the great training robbery Penguin
- Bernstein B (1971) Class, codes and control Routledge and Kegan Paul
- Bibby J (1977) The general linear model: a cautionary tale in O'Muircheartaigh C and Payne C (Eds) The analysis of survey data Vol. 2 Wiley
- Billewicz W Z (1974) Some remarks on experimental intervention studies in relation to malnutrition and its consequences in Cravioto J et al (Eds) Early malnutrition and mental development Almqvist and Wiksell
- Birch H G and Belmont L (1964) Auditory-visual integration in normal and retarded readers Am J Orthopsychiatry 34 852-861
- Bissell J (1973) The cognitive effects of preschool programs for disadvantaged children in Frost J L (Ed) Revisiting early childhood education Holt, Rinehart and Winston
- Bissell J and Smith M S (1970) Letter to Editor Harvard Edul Rev 40 316
- Blackstone T (1975) Education and day care for young children in need: the American experience Centre for Studies in Social Policy
- Blackstone T and Weinreich-Haste H (1980) Why are there so few women scientists and engineers New Society 21 February 1980
- Blalock H M (1964) Causal inferences in non-experimental research W W Norton
- Blalock H M (1971) Causal models in the social sciences Macmillan
- Blalock H M (1972) Theory building and causal inferences in Blalock H M and Blalock A B (Eds) Methodology in social research McGraw-Hill
- Blank M (1972) The wrong response: is it to be ignored, prevented or treated in Parker R K (Ed) The preschool in action Allyn and Bacon
- Blank M and Solomon F (1968) A tutorial program to develop abstract language in socially disadvantaged children C Dev 39 379-389
- Block J, Block J and Harrington D (1975) Comment on the Kagan-Messer reply Dev Psy 11 239-252
- Bloom B (1964) Stability and change in human characteristics John Wiley

- Blum M L and Naylor J C (1968) The interview and other selection devices in Blum M L and Naylor J C (Eds) Industrial Psychology Harper and Row
- Bock R D (1975) Multivariate statistical methods in behavioural research McGraw-Hill
- Bock R D (1975) A multivariate correction for attenuation Biometrika **62** 673-678
- Boehm A (1970) The Boehm Test of Basic Concepts The Psychological Corporation
- Bohrnstedt G (1969) Observations on the measurement of change in Borgatta E F (Ed) Sociological methodology Jossey-Bass
- Bond G L and Dykstra R (1966-7) The cooperative research program in first grade reading instruction Rdg Rsh Quarterly **2** 4 1-142
- Boneau C A (1960) The effects of violation of assumptions underlying the t-test Psy Bulletin **57** 49-64
- Boudon R (1968) A new look at correlation analysis in Blalock H M and Blalock A B (Eds) Methodology in social research McGraw-Hill
- Bougere M (1969) Selected factors in oral language related to first-grade reading achievement Rdg Rsh Quarterly **5** 31-58
- Bowlby J (1953) Child care and the growth of love Pelican
- Bowlby J (1969) Attachment and loss Hogarth
- Boyle R P (1970) Path analysis and ordinal data Am J Sociol **75** 461-480
- Brainerd C (1974) Training and transfer of transitivity, conservation and class inclusion of length C Dev **45** 324-334
- Brandt R (1973) Observational methodology for evaluation of early childhood programs J Rsh and Dev in Edu **6** 3 94-109
- Brimer M A and Dunn L L (1962) The English Picture Vocabulary Test Edul Evaluation Enterprises
- Brimer M A and Dunn L L (1973) The English Picture Vocabulary Test, full range Edul Evaluation Enterprises
- Broman S H et al (1975) Preschool IQ: prenatal and early developmental correlates Lawrence Erlbaum
- Bronfenbrenner U (1974a) The two worlds of childhood Penguin
- Bronfenbrenner U (1974b) Is early intervention effective? U S Dept of Health, Edu and Welfare Nr. 74-25
- Bronfenbrenner U (1974c) Developmental research, public policy and the ecology of childhood C Dev **45** 1-5
- Bronfenbrenner U (1979) The ecology of human development: experiments by nature and by design Harvard U P
- Brown F (1976) Principles of educational and psychological testing Holt, Rinehart and Winston
- Brown P and Payne C (1975) Election night forecasting J Royal Stat Socy, Series A **138** 463-498
- Bruner J (1966) Towards a theory of instruction W W Norton
- Bruner J (1971) The relevance of education Allen and Unwin
- Bruner J (1975) The ontogenesis of speech acts J C Lang **2** 1-19
- Bruner J (1980) Under Five in Britain: Oxford Preschool Research Project Grant McIntyre
- Bruner J, Olver R and Greenfield P (1967) Studies in cognitive growth John Wiley
- Bryant B, Harris M and Newton D (1980) Children and Minders Grant McIntyre
- Bryant P (1972) The understanding of invariance by very young children Canadian J Psy **26** 78-96
- Buffery A W H and Gray J A (1972) Sex differences in the development of spatial and linguistic skills in Ounsted C and Taylor D C (Eds) Gender differences, their ontogeny and significance Churchill-Livingstone
- Bühler C (1943) From birth to maturity Kegan Paul, Trench, Trubner
- Bullock Report (1975) A language for life H.M.S.O.
- Burroughs G E R (1971) Design and analysis in educational research U Birmingham
- Busch J (1978) Television's effect on reading: a case study Phi Delta Kappan **59** 10 668-671
- Butcher H J (1968) Human intelligence - its nature and assessment Methuen
- Butts R P and Cremin L A (1953) The history of education in the American culture Holt, Rinehart and Winston

- Cain G G and Watts H W (1970) Problems in making policy inferences from the Coleman Report Am Sociol Rev 35 228-252
- Cairns E (1977) The reliability of the Matching Familiar Figures Test Br J Edu Psy 47 197-198
- Caldwell B (1974) The malnourishing environment in Cravioto J et al (Eds) Early malnutrition and mental development Almqvist and Wiksell
- Caldwell B (1976) Patterns of upbringing and child care Expert Committee on Children's Mental Health and Psychological Development, Geneva, November 1976
- Caldwell B, Heider J and Kaplan B (1966) The inventory of home stimulation Am Psy Ass, 1966
- Campbell D (1960) Recommendations for APA test standards regarding construct, trait or discriminant validity Am Psychologist 15 546-553
- Campbell D and Erlebacher A (1970) How regression artefacts in quasi-experimental evaluations can mistakenly make compensatory education look harmful in Hellmuth J (Ed) Disadvantaged Child, Vol. 3 Brunner/Mazel
- Campbell D and Stanley J (1963) Experimental and quasi-experimental designs in research on teaching in Gage N L (Ed) Handbook of Research on Teaching Rand McNally
- Cane B and Smithers J (1971) The roots of reading: a study of 12 infant schools in deprived areas National Foundation for Edu Rsh
- Cannell C F and Kahn R L (1968) Interviewing in Lindzey G and Aronson E (Eds) The handbook of social psychology, Vol. 2 Addison-Wesley
- Central Policy Review Staff (1978) Services for young children with working mothers H.M.S.O.
- Chall J (1967) Learning to read: the great debate McGraw-Hill
- Chalmers C (1979) Regression generator Birkbeck College, U London
- Chamberlain R et al (1975) British Births 1970: Vol. 1, The First Week of Life Heinemann
- Chazan M (1975a) The use of tests in the evaluation of pre-school educational programmes in Council of Europe (Eds) Problems in the evaluation of pre-school education Council of Europe Strasbourg
- Chazan M (1975b) Evaluation of pre-school education: research in the United Kingdom in Council of Europe (Eds) Problems in the evaluation of pre-school education Council of Europe Strasbourg
- Chazan M (1978) International research in early childhood education National Foundation for Edu Rsh
- Chazan M and Cox T (1976) Language programmes for disadvantaged children in Varma V P and Williams P (Eds) Piaget, Psychology and Education Hodder and Stoughton
- Chazan M, Laing A and Jackson S (1971) Just before school Blackwell
- Chazan M and Williams P (Eds) (1978) Deprivation and the infant school Schools Council/Basil Blackwell
- Cicirelli V G et al (1969) The impact of Head Start: an evaluation of the effects of Head Start on children's cognitive and affective development Westinghouse Learning Corporation/Ohio University/U S Dept of Commerce
- Clark M (1976) Young fluent readers Heinemann
- Clarke A M and Clarke A D B (1976) Early experience: myth and evidence Open Books
- Clarke W D and Mason J (1969) A programmed introduction to number conservation DES/SSBC
- Clay M (1972) The early detection of reading difficulties: a diagnostic survey Heinemann
- Clements M A (1976) The numeracy breakdown Australian Mathematics Teacher 32 2 45-50
- Clifford C and Wattenberg W W (1962) Relationship of the self-concept to beginning achievement in reading Report CRP-377 Wayne State U Detroit
- Cochrane K (1976) Reading readiness tests: a survey of current trends The exceptional child 23 119-127
- Cohen J and Cohen P (1975) Applied multiple regression and correlation analysis for the behavioural sciences Lawrence Erlbaum
- Coleman J (1970) Reply to Cain and Watts (on making policy inferences from the Coleman Report) Am Sociol Rev 35 242-249

- Coleman J (1975) Methods and results in the IEA studies of effects of school on learning Rev Edul Rsh 45 3 335-386
- Coleman J et al (1966) Equality of educational opportunity U S Office of Edu Washington
- Comber L C and Keeves J P (1973) Science education in 19 countries Int Studies in Evaluation Vol. 2 Almqvist and Wiksell Stockholm
- Coniffe D and Stone J (1973) A critical view of ridge regression The Statistician 22 181-187
- Cooley W W and Lohnes P R (1971) Multivariate data analysis John Wiley
- Costner H (1969) Theory, deduction and rules of correspondence Am J Sociol 75 245-263
- Council of Europe (1975) Problems in the evaluation of pre-school education Documentation Centre for Education in Europe Council of Europe Strasbourg
- Cowles M P (1974) N = 35 : a rule of thumb for psychological researchers Perceptual and Motor Skills 38 1135-1138
- Cox A and Rutter M (1976) Diagnostic appraisal and interviewing in Rutter M and Hersov L (Eds) Child psychiatry: modern approaches Blackwell Scientific
- Crano W D et al (1972) Does intelligence cause achievement? A cross-lagged panel analysis J Edul Psy 63 258-275
- Cravioto J and Delicardie E (1972) Environmental correlates of severe clinical malnutrition and language development in survivors from kwashiorkor or marasmus in Nutrition: the Nervous System and Behaviour Scient. Publ. 251 Pan American Health Organisation Washington
- Cravioto J et al (1974) Early malnutrition and mental development Almqvist and Wiksell
- Cronbach L (1961) Essentials of psychological testing Harper and Row
- Cronbach L (1976) On the design of educational measures in De Gruijter D and Van der Kamp L (Eds) Advances in psychological and educational measurement John Wiley
- Cronbach L and Furby L (1970) How should we measure change - or should we? Psyl Bulletin 74 68-80
- Cronbach L, Gleser G et al (1972) The dependability of behavioural measurements John Wiley
- Cronbach L and Meehl P (1955) Construct validity in psychological testing Psyl Bulletin 52 281-302
- Crowe B (1973) The playgroup movement Allen and Unwin
- Cunningham C (1980) Motivating parents Health Visitor training programmes Child Development Project U Bristol
- Cunningham C and Sloper P (1978) Helping your handicapped child Souvenir Press
- Cureton E (1939) The principal compulsions of factor-analysts Harvard Edul Rev 2 287-295
- Daniels J C and Diack H (1958) The standard reading tests Chatto and Windus
- Darlington R (1968) Multiple regression in psychological research and practice Psyl Bulletin 69 161-182
- Davidson B (1977) Nutritional intervention programme operated by medical graduates Personal communication
- Davie R, Butler N and Goldstein H (1972) From birth to seven Longman
- Day M and Parker R (1977) The Preschool in Action Allyn and Bacon
- Deasey D (1978) Education under six Croom Helm
- Deegan J (1976) The consequences of model misspecification in regression analysis Multivariate Behavioural Rsh 1 2 237-248
- De Finetti B (1961) The Bayesian approach to the rejection of outliers in Neyman J (Ed) Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability U California
- De Hirsch K et al (1966) Predicting reading failure Harper
- Dempster A P (1973) Alternatives to least squares in multiple regression in Kabe D and Gupta R (Eds) Multivariate statistical inference North-Holland
- Dempster A P et al (1977) A simulation study of alternatives to Ordinary Least Squares J Am Stat Ass 72 77-106
- Denney D (1974) Relationship of three cognitive style dimensions to elementary reading abilities J Edul Psy 66 702-709

- D.E.S. (1981) Under fives: a programme of research Monograph Under Fives Research Dissemination Group Department of Education and Science
- Detheux M, Leclercq E et al (1974) From compensatory education to mastery learning London Edul Rev 3 3 41-50
- Deutsch M (1964) Facilitating development in the pre-school child Merrill-Palmer Quarterly 10 3 249-263
- De Vries A (1974) The Utrecht language and thought programme Ph D thesis U Utrecht
- Dimitrovsky L and Almy M (1975) Early conservation as a predictor of arithmetic achievement J Psy 91 65-70
- Dobbing J (1980) Nutritional growth restriction and the nervous system Annales Nestlé No. 44
- Doman G (1975) Teach your baby to read Pan
- Donachy W (1976) Parent participation in pre-school education Br J Edul Psy 46 31-39
- Douglas J W B (1948) Maternity in Great Britain Oxford U P
- Douglas J W B (1964) The home and the school MacGibbon and Kee
- Douglas J W B and Ross J M (1964) The later educational progress and emotional adjustment of children who went to nursery schools or classes Edul Rsh 7 1 73-80
- Douglas J W B et al (1968) All our future Peter Davies
- Downing J (1963) Is a 'mental age of six' essential for 'reading readiness'? Edul Rsh 6 1 16-28
- Downing J (1969) Children's concepts of language in learning to read Edul Rsh 12 106-112
- Downing J (1973) Comparative reading Macmillan
- Downing J (1980) The reading riddle Phi Delta Kappan May 1980 61
- Downing J and Thackray D V (1971) Reading readiness U London Press
- Drabman R S and Hammer D (1977) Using incentives in early childhood education in Hom H L and Robinson P A (Eds) Psychological processes in early education Academic
- Draper N and Smith H (1966) Applied regression analysis John Wiley
- Draper N and Van Nostrand R C (1979) Ridge regression and James Stein estimation: review and comments Technometrics 21 4 451-466
- Duncan O (1966) Path analysis: sociological examples Am J Sociol 72 1-16
- Duncan O (1969) Contingencies in constructing causal models in Borgatta E F (Ed) Sociological methodology Jossey-Bass
- Duncan O (1971) Path analysis: sociological examples (modified version of Duncan, 1966) in Blalock H M (Ed) Causal models in the social sciences Macmillan
- Duncan O (1975) Introduction to structural equation models Academic
- Duncan O and Featherman D L (1973) Psychological and cultural factors in the process of occupational achievement in Goldberger A S and Duncan O D (Eds) Structural equation models in the social sciences Seminar Press
- Dukes W (1970) N = 1 in Badia P et al (Eds) Research problems in psychology Addison-Wesley
- Durbin-Smith (1978) Children and TV Durbin-Smith Communications 2 Basil Street London
- Durkin D (1976) Teaching young children to read Allyn and Bacon
- Dwyer C (1974) Influence of children's sex role standards on reading and arithmetic achievement J Edul Psy 66 811-816
- Dweck C (1975) The role of expectations and attributions in the alleviation of learned helplessness J Pers Soc Psy 31 674-685
- Dweck C and Repucci N D (1973) Learned helplessness and reinforcement responsibility in children J Pers Soc Psy 25 109-116
- Eaves L J, Last K et al (1977) A progressive approach to non-additivity and genotype-environmental covariance in the analysis of human differences Br J Mathematical Psy 30 1-42
- Edmonds R et al (1973) A Black response to Christopher Jenck's Inequality and certain other issues in Harvard Educational Review (Eds) Perspectives on Inequality Harvard Edul Rev Reprint Series No. 8

- Edwards A W F (1972) Likelihood Cambridge U P
- Ehrenberg A S C (1976) Eysenck, random samples or science? Bull Br Psy Socy 29 42-44
- Elardo R et al (1975) The relation of infants' home environments to mental test performance from 6 to 36 months: a longitudinal analysis C Dev 46 71-76
- Elkind D (1974) Two approaches to intelligence: Piagetian and psychometric in Sants J and Butcher H J (Eds) Developmental Psychology Penguin
- Elkind D and Deblinger J A (1969) Perceptual training and reading achievement in disadvantaged children C Dev 40 11-19
- Elkonin D B (1973) Methods of teaching reading in Downing J (Ed) Comparative Reading Collier-Macmillan
- Ellenberg J (1976) Testing for a single outlier from a general linear regression Biometrics 32 3 637-645
- Evans R (1972) Swansea evaluation profile for school entrants Schools Council Research and Development Project in Compensatory Education
- Eysenck H (1964) Crime and personality Paladin
- Eysenck H (1973) Performance contracting: the theory that failed Times Edul Supplement 19 January 1973
- Eysenck H (1975) Who needs a random sample? Bull Br Psy Socy 28 195-198
- Eysenck H and Eysenck S B (1969) Personality structure and measurement Routledge and Kegan Paul
- Faden V and Bobko P (1982) Validity shrinkage in ridge regression: a simulation study Edul Psyl Measurement 42 73-83
- Farebrother R W (1975) The minimum mean square error linear estimator and ridge regression Technometrics 17 127-128
- Farnham-Diggory S (1972) Cognitive processes in education Harper and Row
- Farr R (1972) The Clymer-Barrett Prereading Battery in Buross O K (Ed) The Seventh Mental Measurements Yearbook Gryphon Press
- Farran D and Haskins R (1980) Reciprocal influence in the social interactions of mothers and three-year-old children from different socio-economic backgrounds C Dev 51 780-791
- Feiler A and Thomas G (1980) Teaching children to read - time to despecialise Remedial Edu 15 2 61-65
- Ferguson G (1971) Statistical analysis in psychology and education McGraw-Hill
- Ferguson N, Davies P et al (1971) The Plowden Report's recommendations for identifying children in need of extra help Edul Rsh 13 3 210-213
- Ferree M (1976) Working class jobs: housework and paid work as sources of satisfaction Soc Problems 23 4 431-441
- Ferri E and Niblett R (1977) Disadvantaged families and playgroups National Foundation for Edul Rsh
- Finn J (1974) A general model for multivariate analysis Holt, Rinehart and Winston
- Flavell J (1963) The developmental psychology of Jean Piaget Van Nostrand
- Flavell J (1972) An analysis of cognitive-developmental sequences Genetic Psy Monograph 86 279-350
- Floud J, Halsey A H and Martin F M (1957) Social class and educational opportunity Heinemann
- Fogelman K R and Goldstein H (1976) Social factors associated with changes in educational attainment Edul Studies 2 2 95-109
- Fowler W (1975) How adult/child ratios influence development Interchange 6 1 17-31
- Francis H (1974) Social background, speech and learning to read Br J Edul Psy 44 290-299
- Francis H (1980) A question of method in Tizard B et al (Eds) Fifteen thousand hours: a discussion U London Institute of Education
- Fuller W A and Hidiroglou M A (1978) Regression estimation after correcting for attenuation J Am Stat Ass 73 99-104

- Gallagher J, Ramey C et al (1976) Use of longitudinal research in the study of child development in Tjossem T (Ed) Intervention strategies for high risk infants and young children Baltimore U Park Press
- Galtung J (1967) Theory and methods of social research Allen and Unwin
- Gammage P (1974) Socialisation, Schooling and Locus of Control Ph D thesis U Bristol
- Garber H and Heber R (1973) The Milwaukee Project: early intervention as a technique to prevent mental retardation U Connecticut Technical Paper
- Gardner P L (1975) Attitude measurement: a critique of some recent research Educ Rsh 17 2 101-109
- Gates A (1954) Unsolved problems in reading: a Symposium I Elementary English October 1954
- Gaudy E and Spielberger C D (1971) Anxiety and educational achievement John Wiley
- Gehlbach R (1975) Natural and educational play: some preliminary discriminations Interchange 6 3 1-3
- Gewirtz J (1969) Mechanisms of social learning: some roles of stimulation and behaviour in early human development in Goslin D (Ed) Handbook of socialisation theory and research Rand McNally
- Ghikas C (1978) An exploratory investigation into the effects of 'You and Me', a HBC Schools TV programme for four and five-year-olds M Sc thesis U London
- Gibbons D G (1978) A simulation study of some ridge estimators Research Publication GMR-2659 (revised) General Motors Laboratories
- Gibson E (1972) Reading for some purpose in Kavanagh J and Mattingly I (Eds) Language and reading: an interdisciplinary approach Centre for Applied Linguistics
- Ginsburg H (1972) The myth of the deprived child - poor children's intellect and education Prentice-Hall
- Gnanadesikan R (1977) Methods for statistical data analysis of multivariate observations John Wiley
- Gocka E (1973) Alternate tests for comparing independent groups Psyl Reports 32 683-692
- Goldberg A (1971) Discerning a causal pattern among data on voting in Blalock H M (Ed) Causal models in the social sciences Macmillan
- Goldberg M (1968) Evaluation of compensatory programmes in Passow A H (Ed) Developing programmes for the disadvantaged Teachers College Press Columbia U
- Goldberger A (1973) Efficient estimation in overidentified models: an interpretive analysis in Goldberger A S and Duncan O D (Eds) Structural equation models for the social sciences Seminar Press
- Goldberger A and Duncan O (1973) Structural equation models for the social sciences Seminar Press
- Goldstein H (1972) The construction of standards for measurement subject to growth Human Biology 44 255-261
- Goldstein H (1976a) Whose values in educational research? Bias 3 1
- Goldstein H (1976b) Review of Applied multiple and correlation analysis for the behavioural sciences (by Cohen J and Cohen P) J Royal Stat Socy 139 4 549
- Goldstein H (1976c) A brief review of some approaches to the statistical analysis of longitudinal data Seminar S S R C Survey Unit
- Goldstein H (1977) Multiple regression (personal communication) Seminars Institute of Education U London
- Goldstein H (1979) The design and analysis of longitudinal studies Academic
- Goldstein H and Fogelman K (1974) Age standardisation and seasonal effects in mental testing Br J Educ Psy 44 109-115
- Goldstein M and Smith A F M (1974) Ridge type estimators for regression analysis J Royal Stat Socy Series B 36 284-291
- Golub G H et al (1979) Generalised cross-validation as a method for choosing a good ridge parameter Technometrics 21 215-223
- Goodacre E (1967) Reading in infant classes National Foundation for Educ Rsh
- Goodenough F L (1926) Measurement of intelligence by drawings World Book
- Goodman L (1973) Causal analysis of data from panel studies and other kinds of surveys Am J Sociol 78 1135-1191

- Gordon I (1973) A home learning centre approach to early stimulation in  
 Frost J L (Ed) Revisiting early childhood education Holt, Rinehart, Winston
- Gordon I et al (1977) The Florida parent education infant and toddler programs  
 in Day M C and Parker R K (Eds) The preschool in action Allyn and Bacon
- Gordon R (1968) Issues in multiple regression Am J Sociol 73 592-616
- Gorsuch R (1973) Data analysis of correlated independent variables Multi-  
 variate Behavioural Rsh 8 89-108
- Gottman J (1973) N-of-one and N-of-two in psychotherapy Psyl Bull 80 93-105
- Gourlay N (1953) Covariance analysis and its applications in psychological re-  
 search Br J Mathematical Stat Psy 6 25-34
- Gray S and Klaus R A (1972) The early training project: a seventh-year report  
 in Bronfenbrenner U (Ed) Influences on human development Dryden
- Gray S and Wandersman L (1980) The methodology of home-based intervention  
 studies: problems and promising strategies C Dev 51 993-1009
- Green B (1977) Parameter sensitivity in multivariate analysis methods  
J Multivariate Behavioural Rsh 12 263-287
- Green R T and Laxon V (1977) Entering the world of number (Monograph: perso-  
 nal communication)
- Grieger P and Saavedra P (1971) Pygmalion revisited: a loud call for caution  
Interchange 2 4
- Groff P (1975) Reading ability and auditory discrimination: are they related?  
Rdg Teacher 28 742-747
- Guilkey D K and Murphy J L (1975) Directed ridge regression techniques in cases  
 of multicollinearity J Am Stat Ass 70 769-775
- Gulliford R (1969) Backwardness and educational failure National Foundation  
 for Edul Rsh
- Gumbert E and Spring J (1974) The superschool and the superstate - American  
 education in the twentieth century, 1918-1970 John Wiley
- Gunst R F and Mason R L (1977) Biased estimation in regression: an evaluation  
 using mean squared error J Am Stat Ass 72 616-628
- Gutelius M F, Kirsch A D et al (1972) Promising results from a cognitive stimu-  
 lation programme in infancy Clinical Pediatrics 11 10 585-593
- Haavelmo T (1943) The statistical implications of a system of simultaneous  
 equations Econometrica 11 1 1-12
- Halliday M A K (1975) The infant's communication of meaning Seminar Centre  
 for Urban Educational Studies London
- Halsey A H (1972) Educational Priority, Vol. 1 H M S O
- Halsey A H, Heath A F and Ridge J M (1980) Origins and destinations: family,  
 class and education in modern Britain Clarendon Oxford
- Hamill D (1972) Training visual perceptual processes J Learning Disabilities  
 5 39-46
- Hardy J B, Welcher D W et al (1976) Pitfalls in the measurement of intelligence:  
 are standard intelligence tests valid instruments for measuring the intellec-  
 tual potential of urban children? J Psy 94 43-51
- Harman H (1967) Modern factor analysis U Chicago
- Harris C W (1963) Problems in measuring change U Wisconsin Press
- Harris D (1963) Children's drawings a measures of intellectual maturity  
 Harcourt, Brace and World
- Harris L J (1975) Interaction of experiential and neurological factors in the  
 patterning of human abilities: the question of sex differences in 'right  
 hemisphere skills Symposium Society for Research in Child Development Denver
- Harvard Educational Review (1973) Perspectives on inequality Reprint Series 8
- Hauser R (1971) Socioeconomic background and educational performance American  
 Sociological Association
- Hauser R (1973) Disaggregating a social-psychological model of educational  
 attainment in Goldberger A S and Duncan O D (Eds) Structural equation  
 models in the social sciences
- Hauser R and Goldberger A S (1971) The treatment of unobservable variables in  
 path analysis in Costner H L (Ed) Sociological Methodology 1970 Jossey-  
 Bass



- Havlicek L L and Peterson N L (1977) The effect of the violation of assumptions upon significance levels of the Pearson correlation coefficient Psyl Bull 84 2 373-377
- Hays W (1969) Statistics Holt, Rinehart and Winston
- Hays W (1973) Statistics for the Social Sciences Holt, Rinehart and Winston
- Heath A and Clifford P (1980) The seventy thousand hours that Rutter left out Oxford Rev Edu 6 1 3-19
- Heber R and Garber H (1975) The Milwaukee Project: a study of family intervention to prevent cultural familial mental retardation in Friedlander B Z et al (Eds) Exceptional Infant, Vol. 3: Assessment and Intervention Brunner/Mazel
- Hewison J (1980) Home environment and reading attainment: a study of children in a working class community Ph D thesis Institute of Education U London
- Hewison J (1981) Home is where the help is Times Edul Supplement 16 Jan 1981
- Hewison J and Tizard J (1980) Parental involvement and reading attainment Br J Edu Psy 50 209-215
- Hindley C (1960) General considerations in Falkner F (Ed) Modern problems in pediatrics, Vol. 5: Child Development - an international method of study Karger, Basel
- Hindley C (1965) Stability and change in abilities up to five years: group trends J C Psy Psychiatry 6 85-99
- Hindley C (1972) The place of longitudinal methods in the study of development in Minks W W et al (Eds) Determinants of behavioural development Academic
- Hindley C (1977) Problems in ascribing causation Seminar Institute of Education U London
- Hindley C and Owen C (1978) The extent of individual changes in IQ for ages between 6 months and 17 years, in a British longitudinal sample J C Psy Psychiatry 19 329-350
- Hochberg J (1970) Components of literacy: speculations and exploratory research in Levin H and Williams J (Eds) Basic studies in reading Basic Books
- Hocking R R et al (1976) A class of biased estimators in linear regression Technometrics 18 425-437
- Hoerl A E (1962) Application of ridge analysis to regression problems Chemical Engineering Progress 58 54-59
- Hoerl A E and Kennard R W (1970a) Ridge regression: biased estimation for non-orthogonal problems Technometrics 12 55-67
- Hoerl A E and Kennard R W (1970b) Ridge regression: application to non-orthogonal problems Technometrics 69-82
- Holland S (1979) Rutland Street Bernard van Leer Foundation / Pergamon Press
- Holmes D (1973) The independence of letter, word and meaning identification in reading in Smith F (Ed) Psycholinguistics and reading Holt, Rinehart and Winston
- Holt J (1969) Schools are bad places for kids Saturday Evening Post 8 Feb '69
- Horowitz F D and Paden L Y (1973) The effectiveness of educational intervention programmes in Horowitz F D (Ed) Review of child development research, Vol 3 U Chicago Press
- House E (1978) The objectivity, fairness and justice of Federal evaluation as reflected in the Follow Through evaluation Centre for Instructional Rsh and Curriculum Evaluation U Illinois
- House E, Glass G et al (1978) No simple answer: critique of the Follow Through evaluation Harvard Edul Rev 48 128-160
- Hughes M, Mayall B et al (1980) Nurseries now: a fair deal for parents and children Harmondsworth
- Hull C L (1943) Principles of behaviour Appleton-Century-Crofts
- Hunt J McV (1961) Intelligence and experience Ronald Press
- Hunt J McV (1971) Parent and child centres: their basis in the behavioural sciences Am J Orthopsychiatry 41 1 37-63
- Hutchison D, Prosser H and Wedge P (1979) The prediction of educational failure Edu Studies 5 73-82
- Hutt C (1972) Males and females Penguin
- Hutt C (1974) Sex: what's the difference? Science 62 898 405-407

- Hutt C (1979) Play in the under-fives in Howells J G (Ed) Modern perspectives in the psychiatry of infancy Brunner/Mazel
- Hutt S J and Hutt C (1970) Direct observation and measurement of behaviour Charles Thomas
- Hyman H et al (1954) Interviewing in social research U Chicago
- Igaga J M and Versey J (1978) Cultural differences in rhythmic performance  
Psy of Music 6 1 61-65
- Illich I (1974) Schooling: the ritual of progress in Lister I (Ed) Deschooling Cambridge U P
- Jencks C et al (1972) Inequality: a reassessment of the effect of family and schooling in America Basic Books
- Jensen A (1969) How much can we boost IQ and scholastic achievement?  
Harvard Edul Rev 39 1-123
- Jensen A (1977) Cumulative deficit in IQ of Blacks in the rural South  
Dev Psy 13 184-191
- Johnson D (1975) The development of a programme for parent-child education among Mexican-Americans in Texas in Friedlander B Z et al (Eds) Exceptional Infant, Vol. 3: assessment and intervention Brunner/Mazel
- Johnson S and Bolstad O (1973) Methodological issues in naturalistic observation: some problems and solutions for field research in Hamerlynck L A et al (Eds) Behavioural change: methodology, concepts and practice Research Press
- Johnston J (1963) Econometric methods McGraw-Hill
- Johnston J (1972) Econometric methods, 2nd Edition McGraw-Hill
- Jöreskog K (1976) Structural equation models in the social sciences: specification, estimation and testing Rsh Rep 76-9 Uppsala
- Jöreskog K and Sörbom D (1976) Statistical models and methods for test-retest situations in De Gruijter D and Van der Kamp L (Eds) Advances in psychological and educational measurement John Wiley
- Kagan J (1965) Reflection-impulsivity and reading ability in primary grade children C Dev 36 609-628
- Kagan J (1969) Inadequate evidence and illogical conclusions Harvard Edul Rev 39 274-277
- Kagan J (1971) Developmental studies in reflection and analysis in Open University (Eds) Personality growth and learning Longman / Open U P
- Kagan J and Moss H A (1962) Birth to maturity: a study in psychological development John Wiley
- Kamii C and Elliott D L (1971) Evaluation of evaluators Educational Leadership 28 8 827-831
- Karnes M B ((1973) Evaluation and implications of research with young handicapped and low-income children in Stanley J C (Ed) Compensatory education for children ages two to eight John Hopkins
- Karnes M B, Studley W M et al (1968) An approach for working with mothers of disadvantaged pre-school children Merrill-Palmer Quarterly 14 174-184
- Karnes M B and Zehrbach R R (1977) Educational intervention at home in Day M C and Parker R K (Eds) The preschool in action Allyn and Bacon
- Keddie N (1973) (Ed) Tinker, tailor - the myth of cultural deprivation Penguin
- Kellaghan T (1977) The evaluation of an intervention project for disadvantaged children National Foundation for Edul Rsh
- Kelsall R K and Kelsall H (1971) Social disadvantage and educational opportunity Holt, Rinehart and Winston
- Kendall M (1957) A course in multivariate analysis Charles Griffin
- Kendall M (1975) Multivariate analysis Charles Griffin
- Kendall M and Stuart A (1963) The advanced theory of statistics, Vol. 2 Hafner Press
- Keniston K and the Carnegie Council on Children (1977) All our children: the American family under pressure Harcourt, Brace, Jevanovitch
- Kerlinger F (1973) Foundations of behavioural research Holt, Rinehart, Winston

- Kerlinger F and Pedhazur E (1973) Multiple regression in behavioural research Holt, Rinehart and Winston
- Kessen W (1960) Research design in the study of developmental problems in Mussen P H (Ed) Handbook of research methods in child development John Wiley
- Kim J (1975) Multivariate analysis of ordinal variables Am J Sociol 81 2 261-298
- King W H (1969) Statistics in education Macmillan
- Kish L (1975) Representation, randomisation and control in Blalock H et al (Eds) Quantitative Sociology Academic
- Kitay P (1972) The Bender-Gestalt in Buros O K (Ed) The seventh mental measurements yearbook Gryphon
- Klein L (1974) A textbook of econometrics Prentice-Hall
- Koppitz E (1960) The Bender Gestalt test for children: a normative study J Clinical Psy 16 432-435
- Koppitz E (1964) The Bender Gestalt test for young children Grune and Stratton
- Koppitz E (1975) Bender Gestalt test for young children, Vol.2: research and application Grune and Stratton
- Kosc L (1974) Developmental dyscalculia J Learning Disabilities 7 3 164-177
- Koutsoyiannis A (1977) Theory of Econometrics Macmillan
- Krutelskii V A (1976) The psychology of mathematical abilities in school children U Chicago Press
- Labov W (1972) The study of language and its social context in Labov W (Ed) Sociolinguistic patterns U Pennsylvania Press
- Laing A (1973) Structure in preschool education in Chazan M (Ed) Education in the early years Swansea U College, Faculty of Education
- Lambie D, Bond J and Weikart D (1974) Home teaching with mothers and infants The Ypsilanti-Carnegie Infant Education Project
- Land, K (1969) Principles of path analysis in Borgatta E F (Ed) Sociological methodology 1969 Jossey-Bass
- Landsdown R (1978) Retardation in mathematics: a consideration of multifactorial determination J C Psy Psychiatry 19 181-185
- Lashley K S (1967) The problem of serial order in behaviour in Jeffries L A (Ed) Cerebral mechanisms in behaviour Hafner
- Lawley D N and Maxwell A E (1973) Regression and factor analysis Biometrika 60 331-338
- Lawrence D (1970) The effects of counselling on retarded readers Eduol Rsh 13 119-124
- Lawrence D (1981) The development of a self-esteem questionnaire Br J Eduol Psy 51 245-281
- Lazar I, Hubbell V R et al (1977) The persistence of preschool effects Community Service Laboratory Cornell U
- Lazar I and Darlington R B (1978) (Eds) Lasting effects after preschool Education Commission of the States
- Leach P (1979) Who cares? Penguin
- Leitenberg H (1973) The use of single case methodology J Abnormal Psy 82 87-101
- Lesser H (1977) Television and the preschool child Academic
- Levenstein P (1977) The mother-child home program in Day M C and Parker R K (Eds) The preschool in action Allyn and Bacon
- Levin H (1975) Cost-effectiveness analysis in evaluation research in Gutten-tag M and Struening E (Eds) Handbook of Evaluation Research, Vol.2 Sage
- Lindeman R (1967) Educational measurement Scott, Foresman
- Lindley D V (1965) Introduction to probability and statistics from a Bayesian viewpoint Cambridge U P
- Lindley D V and Smith A F M (1972) Bayes estimators for the linear model J Royal Stat Socy Series B 34 1-41
- Lines R and Widlake P (1971) A structured number conservation programme in Birmingham Educational Priority Areas Action Research Project Report to the S S R C July 1971
- Lister I (1971) Should schools survive? Locum International Conference 'Towards a freer school' West Germany

- Lister I (1974) Deschooling Cambridge U P
- Little A (1978) Education policies for multi-racial areas Goldsmith's College U London
- Lohnes P R (1966) Measuring adolescent personality American Institute for Research Pittsburg
- Lord F (1956) The measurement of growth EduL Psyl Measurement 16 421-437
- Lord F (1958) Further problems in the measurement of growth EduL Psyl Measurement 18 435-451
- Lord F (1963) Elementary models for measuring change in Harris C W (Ed) Problems in measuring change U Wisconsin Press
- Lovell K et al (1962) An experimental study of the growth of some logical structures Br J EduL Psy 53 175-188
- Lovell K (1971a) The growth of understanding in mathematics: kindergarten through grade three Holt, Rinehart and Winston
- Lovell K (1971b) The development of some mathematical ideas in elementary school pupils in Roskopf M F et al (Eds) Piagetian cognitive-developmental research and mathematical education National Council of Teachers of Mathematics
- Lovell K (1972) English Picture Vocabulary Test in Buros O K (Ed) The Seventh Mental Measurements Yearbook Gryphon
- Lovell K (1975) Summary and implications in Steffe L (Ed) Research on mathematical thinking of young children National Council of Teachers of Mathematics
- Lunzer E A et al (1976) The effectiveness of measures of operatively, language and short-term memory in the prediction of reading and mathematical understanding Br J EduL Psy 46 295-305
- McCall R (1970) The use of multivariate procedures in developmental psychology in Mussen P (Ed) Manual of child psychology, Vol.1 John Wiley
- Maccoby E E and Maccoby N (1954) The interview: a tool of social science in Lindzey G (Ed) Handbook of social psychology, Vol.1 Addison-Wesley
- McDonald G C and Ayers J A (1974) Some applications of the "Chernoff faces" Research Publication GMR-1555 General Motors Research Laboratories
- McDonald G C and Galarneau D I (1975) A Monte Carlo evaluation of some ridge type estimators J Am Stat Ass 70 407-416
- McDonald G C and Schwing R (1973) Instabilities of regression estimates relating air pollution to mortality Technometrics 15 463-481
- Macdonald K I (1977) Path analysis in O'Muircheartaigh C and Payne C (Eds) The analysis of survey data, Vol.2: model fitting John Wiley
- Mackay D and Simo J (1976) Help your child to read and write, and more Penguin
- Mackay D, Thompson B and Schaub P (1970) Breakthrough to literacy: teacher's manual Schools Council
- McKay H, Sinisterra L et al (1978) Improving cognitive ability in chronically deprived children Science 200 270-278
- McKennell A C (1977) Attitude scale construction in O'Muircheartaigh C and Payne C (Eds) The analysis of survey data, Vol.1: exploring data structures John Wiley
- McMillan J (1973) Help your baby to read Elliot Right Way
- McNemar Q (1940) Sampling in psychological research Psyl Bull 37 331-365
- McNemar Q (1945) Note on Wellman's re-analysis of IQ changes of orphanage preschool children J Genetic Psy 67 215-219
- McNemar Q (1969) Psychological statistics John Wiley
- Madrid-Jimenez R (1977) Effects of day care: assessment of 12 behavioural traits M Sc thesis Institute of Education U London
- Maliphand R et al (1974) Acquiring skill in reading: a review of experimental research J C Psy Psychiatry 15 175-185
- Marquardt D W and Snee R d (1975) Ridge regression in practice Am Statistician 29 3-20
- Maslow A H (1943) A theory of human motivation Psyl Rev 50 370-396
- Matthews G and Matthews J (1974-77) Early mathematical experiences Schools Council

- Mattingly I (1972) Reading, the linguistic process, and linguistic awareness in Kavanagh J and Mattingly I (Eds) Language by ear and by eye M.I.T. Press
- Maxwell A E (1972) The WPPSI: a marked discrepancy in the correlations of the sub-tests for good and poor readers Br J Mathematical Stat Psy 28 51-62
- Maxwell A E (1975) Limitations on the use of the multiple linear regression model Br J Mathematical Stat Psy 28 51-62
- Maxwell A E (1977) Multivariate analysis in behavioural research Chapman, Hall
- Mayall B and Petrie P (1981) Day care for children under two years old in Under Fives Research Dissemination Group (Eds) Under Fives: a programme of research Dept of Edu and Science
- Meehl P E (1970) Theory testing in psychology and physics: a methodological paradox in Morrison D H and Henkel R F (Eds) The significance controversy Butterworth
- Miller L B and Dyer J L (1975) Four preschool programs: their dimensions and effects Monographs Socy Rsh C Dev 40 5-6 Serial No. 162
- Mischel W (1971) Introduction to personality Holt, Rinehart and Winston
- Mischel W (1973) Empirical dilemmas of psychological approaches J Abnormal Psy 82 335-344
- Modgil S (1974) Piagetian research: a handbook of recent studies National Foundation Edul Rsh
- Moore T, Hindley C and Falkner F (1954) A longitudinal research in child development and some of its problems Br Medical J 2 1132-1137
- Morris A G et al (1976) Educational intervention for preschool children in a pediatric clinic Pediatrics 57 5 765-768
- Morrison D H and Henkel R F (1970) (Eds) The significance controversy Butterworth
- Mortimore J and Blackstone T (1982) Disadvantage and education Heinemann
- Mosbaeck E J and Wold H O (1970) Interdependent systems: structure and estimation North-Holland
- Moseley C and Moseley D (1977) Language and reading among under-achievers National Foundation Edul Rsh
- Mosteller F and Moynihan D (1972) (Eds) On equality of educational opportunity Random House
- Mosteller F and Tukey J (1977) Data analysis and regression Addison-Wesley
- Mpiangu B D and Gentile J R (1975) Is conservation of number a necessary condition for mathematical understanding? J Rsh Mathematics Edu May 1975 179-192
- Namboordiri N, Carter L and Blalock H (1975) Applied multivariate analysis and experimental designs McGraw-Hill
- Narrol H and Bachor D G (1975) An introduction to Feuerstein's approach to assessing and developing cognitive potential Interchange 6 1 2-16
- Neligan G, Prudham D and Steiner H (1974) The formative years: birth, family and development in Newcastle-upon-Tyne Oxford U P
- New Society (1981) What's happening to family life New Society 30 April 1981
- Newsom Report (1963) Half our future H.M.S.O.
- Newson J (1974) Towards a theory of infant understanding Bull Br Psy Socy 27 251-257
- Newson J and Newson E (1963) Infant care in an urban community Allen and Unwin
- Newson J and Newson E (1968) Four years old in an urban community Allen and Unwin
- Newson J and Newson E (1976a) Parental roles and social contexts in Shipman M (Ed) The organisation and impact of social research Routledge and Kegan Paul
- Newson J and Newson E (1976b) Seven years old in the home environment Allen and Unwin
- Noonan R and Wold H (1977) Nipals path modelling with latent variables Scandinavian J Edul Rsh 21 1 33-61
- Novick M R et al (1972) Estimating multiple regressions in n groups: a cross-validation study Br J Mathematical Stat Psy 25 33-50
- Nunnally J (1970) The place of statistics in psychology in Badia P et al (Eds) Research problems in psychology Addison-Wesley

- Obenchain R L (1975) Ridge analysis following a preliminary test of the shrunk-en hypothesis Technometrics 17 431-441
- Olson D (1970) Cognitive development: the child's concept of diagonality Academic
- O'Muircheartaigh C and Payne C (1977) (Eds) The analysis of survey data. Vols.1 and 2 John Wiley
- Osterreith P, De Coster W et al (1977) Improving education for disadvantaged children Bernard van Leer Foundation / Pergamon
- Overton E and Eversley D (1977) Falling fertility: the effect on the 0-5 age group of the changing demographic structure Centre for Studies in Social Policy
- Owen G (1977) Health visiting Ballière Tindall
- Pantin G (1979) A mole cricket called Servol Bernard van Leer Foundation / Pergamon
- Parry H and Crossley H (1950) Validity of responses to survey questions Public Opinion Quarterly 14 61-80
- Pasmanick B and Knobloch H (1960) Brain damage and reproductive casualty Am J Orthopsychiatry 30 298-305
- Peaker G (1967a) Sampling in Husein T (Ed) International study of achievement in mathematics, Vol.1 John Wiley
- Peaker G (1967b) The regression analyses of the national survey in Plowden Report (Eds) Children and their primary schools, Vol.2, Research and Surveys H.M.S.O.
- Perney J et al (1978) Television viewing: its relationship to early school achievement National College of Education Evanston Ill
- Phares J (1976) Changes in beliefs about locus of control General Learning P
- Piaget J and Inhelder B (1969) The psychology of the child Routledge and Kegan Paul
- Pidgeon D and Yates A (1968) Introduction to educational measurement Routledge and Kegan Paul
- Plowden B (1982) We didn't know then what we know now Times Educl Supplement 2 April 1982
- Plowden Report (1967) Children and their primary schools: Vol.1 Central Advisory Council for Education H M S O
- Plutchik R (1974) Foundations of experimental research Harper and Row
- Pond C and McNay M (1980) Low pay and family poverty Study Commission for the Family 231 Baker Street London NW 1
- Popper K (1976) Unended quest Fontana / Collins
- Posner M et al (1970) Component processes in reading: a performance analysis in Kavanagh J and Mattingly I (Eds) Language by ear and by eye M.I.T. Press
- Poulton G A and James T (1975) Preschool learning in the community - strategies for change Routledge and Kegan Paul
- Pressman L (1982) Sesame Street revisited Contact (ILEA) 11 24 3 Dec 1982
- Price B (1977) Ridge regression: application to non-experimental data Psy Bull 84 759-766
- Pumfrey P (1976) Reading: tests and assessment techniques Hodder and Stoughton / U K Reading Association
- Quigley H (1971) Nursery teachers' reactions to the Peabody Language Development Kit Br J Educl Psy 41 155-162
- Ramey C and Gallagher J (1975) The nature of cultural deprivation: theoretical issues and suggested research strategies North Carolina J Mental Health 7 41-47
- Ramey C et al (1977a) Assessing the intellectual consequences of early intervention with high risk infants Am J Mental Deficiency 81 4 318-324
- Ramey C et al (1977b) An introduction to the Carolina Abecedarian Project in Caldwell B and Stedman D (Eds) Infant education for handicapped children Walker and Co. New York

- Raven J (1980) Parents, teachers and children: a study of an educational home visiting scheme Hodder and Stoughton
- Raven J et al (1978) Priorities in adult education: an evaluation of the Lothian Region educational home visiting programme Scottish Council for Rsh in Edu
- Reid J (1966) Learning to think about reading Edu Rsh 9 56-62
- Reid J (1972) (Ed) Reading problems and practices Ward Lock
- Ribich T (1968) Education and poverty Brookings Institution
- Richards M (1974) (Ed) The integration of a child into the social world Cambridge U P
- Richardson K and Spears D (1972) Race, culture and intelligence Penguin
- Richardson S A, Birch H G et al (1972) The behaviour of children in school who were severely malnourished in the first two years of life J Health Soc Behav 13 276-284
- Richardson S A, Dohrenwend B and Klein D (1965) Interviewing: its form and functions Basic Books
- Rose N (1973) Ten therapeutic playgroups National Society for the Prevention of Cruelty to Children
- Rosen H (1972) Language and class: a critical look at the theories of Basil Bernstein Falling Wall Press
- Rosen B C and D'Andrade R (1972) The psycho-social origins of achievement motivation in Bronfenbrenner U (Ed) Influences on human development Dryden
- Rosenfield A (1975) Psycholinguistic variables in the prediction of reading ability J Genetic Psy 126 309-310
- Rosenthal M (1973) Attachment and mother-infant interaction: some research impasse and a suggested change in orientation J C Psy-Psychiatry 14 201-207
- Rosenthal R (1976) Experimenter effects in behavioural research John Wiley
- Rosenthal R and Jacobson L (1968) Pygmalion in the classroom: teacher expectation and pupils' intellectual development
- Rossetti Ferreira M (1978) Malnutrition and mother-infant asynchrony: slow mental development International J Behavioural Dev 1 207-219
- Rothenberg J (1975) Cost-benefit analysis: a methodological exposition in Guttentag M and Struening E (Eds) Handbook of evaluation research, Vol.2 Sage
- Rozeboom W (1979) Ridge regression: bonanza or beguilement? Psyl Bull 86 2 242-249
- Runkel P and McGrath J (1972) Research on human behaviour Holt, Rinehart and Winston
- Rutter M (1972) Maternal deprivation reassessed Penguin
- Rutter M and Madge N (1976) Cycles of disadvantage Heinemann
- Rutter M, Maughan B, Mortimore P et al (1979) Fifteen thousand hours Open Books
- Rutter M and Yule W (1972) Reading retardation and anti-social behaviour - the nature of the association in Reid J (Ed) Reading problems and practices Ward Lock
- Sanderson A E (1963) The idea of reading readiness: a re-examination Edu Rsh 6 1 3-9
- Sants J and Butcher H (1975) Developmental Psychology Penguin
- Sattler J (1974) Assessment of children's intelligence Saunders
- Schaefer E (1972) Parents as educators: evidence from cross-sectional, longitudinal and intervention research in Hartup W (Ed) The young child: reviews of research, Vol.2 National Ass Edu Young Children
- Schaffer H R (1971) The growth of sociability Penguin
- Schaffer H R (1974) Early social behaviour and the study of reciprocity Bull Br Psy Socy 27 209-216
- Schae K W (1970) A reinterpretation of age-related changes in cognitive structure and functioning in Goulet E and Baltes P (Eds) Life-span developmental psychology: research and theory Academic
- Schae K W (1972) Can the longitudinal method be applied to psychological studies of human development in Munks P et al (Eds) Determinants of human behaviour Academic

- Schmidt F (1971) The relative efficiency of regression and simple unit predictor weights in applied differential psychology Educ Psychol Measurement 31 699-714
- Schmidt F (1972) The reliability of differences between linear regression weights in applied differential psychology Educ Psychol Measurement 32 879-886
- Schönemann P and Steiger J (1976) Regression component analysis Br J Mathematical Stat Psych 29 175-189
- Schramm W et al (1961) Television in the lives of our children Stanford U P
- Schultz C and Pomerantz M (1974) Some problems in the application of achievement motivation to education: the assessment of motivation to succeed and probability of success J Educ Psych 66 599-608
- Seaver W B (1973) Effects of naturally induced expectancies J Pers Soc Psych 28 333-342
- Seligman M (1975) Helplessness: on depression, development and death W H Freeman
- Sharot T (1976) The generalised jack-knife: finite samples and sub-sample sizes J Am Stat Ass 71 451-453
- Shearer D (1980) The Portage Project - the first ten years Portage Wisconsin
- Shipman V (1973) Disadvantaged children and their first school experiences: ETS-Headstart longitudinal study John Hopkins
- Sigel I (1974) When do we know what a child knows? Human Dev 17 201-217
- Sigel I and Hooper F (1968) Logical thinking in children: research based on Piaget's theory Holt, Rinehart and Winston
- Silverstein A B (1970a) Reappraisal of the validity of a short form of Wechsler's scales Psychol Reps 26 559-561
- Silverstein A B (1970b) Reappraisal of the validity of WAIS, WISC and WPPSI short forms J Consulting Clinical Psych 34 1 12-14
- Simon H (1971) Spurious correlation: a causal interpretation in Blalock H (Ed) Causal models in the social sciences Macmillan
- Sinclair H (1971) Number and measurement in Roszkopf M et al (Eds) Piagetian cognitive-development research and mathematical education National Council of Teachers of Mathematics
- Skatkin L (1971) Use of set theory concepts as the basis for teaching mathematics Soviet Education 13 72-79
- Skeels H (1966) Adult status of children with contrasting early life experiences: a follow-up study Monog Socy Res Child Dev 31 3 No.105
- Slater M (1978) 'Significance tests' or Bayesian analysis? An elementary example Bias 5 2 193-217
- Smilansky M (1979) Priorities in adult education: pre-school, evidence and conclusions World Bank Staff working paper No.323
- Smilansky S et al (1976) Mental development of infants from two ethnic groups Henrietta Szold Institute Jerusalem
- Smith F (1971) Understanding reading Holt, Rinehart and Winston
- Smith F (1973) Psycholinguistics and reading Holt, Rinehart and Winston
- Smith F (1975) The relation between spoken and written language in Lenneberg E H and Lenneberg E (Eds) Foundations of language development, Vol.2 Academic
- Smith F (1978) Reading Cambridge U P
- Smith G (1975) Educational Priority, Vol.4: The West Riding project D.E.S./S.S.R.C. H.M.S.O.
- Smith J M (1972) Interviewing in market and social research Routledge and Kegan Paul
- Smith K J (1972) Clymer-Barrett Prereading Battery in Baros O (Ed) The Seventh Mental Measurements Yearbook The Gryphon Press
- Smith M (1968) School and home: focus on achievement in Passow A (Ed) Developing programmes for the educationally disadvantaged Teacher's College Columbia
- Smith T (1980) Parents and preschool Grant McIntyre
- Snedecor G and Cochran W (1967) Statistical methods Iowa State U P
- Southgate V (1959) Southgate Group Reading Tests U London Press
- Southgate V (1972) Beginning reading U London Press



- Sowell T (1976) Patterns of black excellence The Public Interest 43 26-58
- Sowell T (1980) Assumptions versus history in ethnic education Monograph  
The Urban Institute Washington
- Spearman C (1904) The proof and measurement of association between two things  
Am J Psy 15 72-101
- Spencer M (1982) A topic on the boil Times Edul Supplement 30 April 1982
- Stahl D (1973) Relationship between Piaget-based criterion tasks and first  
grade arithmetic achievement Dissertation Abstracts International 34 5A 2181
- Steffe L and Smock C (1975) On a model for learning and teaching mathematics  
in Steffe L (Ed) Research on mathematical thinking of young children  
National Council of Teachers of Mathematics
- Stevens G and Orem R (1968) The case for early reading Warren Green
- Stevenson H, Parker T and Wilkinson A (1976) Predictive value of teachers'  
ratings of young children J Edul Psy 68 507-517
- Stone M (1974) Cross-validatory choice and assessment of statistical predic-  
tion J Royal Stat Socy Series B 36 111-147
- Stott D and Sharp J (1976) Effectiveness motivation in pre-school children  
Edul Rsh 18 2 117-125
- Strauss S (1972) Inducing cognitive development and learning: a review of  
short-term training experiments. 1. The organismic-developmental approach  
Cognition 1 329-357
- Streissguth A and Bee H (1972) Mother-child interactions and cognitive develop-  
ment of children in Hartup W (Ed) The young child: reviews of research,  
Vol.2 National Ass Edu Young Children
- Stukat K G (1975) Introduction in Council of Europe (Eds) Problems in the  
evaluation of pre-school education Documentation Centre for Education in  
Europe Strasbourg
- Sullivan J (1974) Multiple indicators: some criteria of selection in  
Blalock H (Ed) Measurement in the social sciences Macmillan
- Summerfield G (1970) (Ed) The first book Penguin
- Swan D (1975) Reading and intelligence - an integrating theory in Moyle D  
(Ed) Reading: what of the future? Ward Lock
- Swenson E (1973) Teaching mathematics to children Macmillan
- Sylva K et al (1980) Child-watching at playgroup and nursery school Grant  
McIntyre
- Tanner J (1962) Growth at adolescence Blackwell
- Tanner J (1963) The regulation of human growth C Dev 34 817-847
- Tanner J and Inhelder B (1959) Discussions of child development Tavistock
- Taruntaeva T (1971) "Mathematics" in kindergarten (towards a new curriculum)  
Soviet Edu 13 97-107
- Taylor M (1981) Caught between: a review of research into the education of  
pupils of West Indian origin National Foundation Edul Rsh
- Taylor P et al (1972) A study of nursery education Schools Council Working  
Paper 41 Evans/Methuen
- Tew B (1976) Some doubts about the Frestig test of visual perception Remedial  
Edu 11 1 32-35
- Thackray D (1965) The relationship between reading readiness and reading  
progress Br J Edul Psy 35 252-254
- Theil H (1971) Principles of econometrics John Wiley
- Thisted R (1976) Ridge regression, minimax estimation and empirical Bayes  
methods Technical Rep 28 Division of Biostatistics Stanford U
- Thomas A, Chess S and Birch H (1968) Temperament and behaviour disorders in  
children U London Press
- Thorndike R (1973) Reading comprehension education in 15 countries Inter-  
national Studies in Evaluation Vol.3 Almqvist and Wiksell Stockholm
- Times, The (1971) 'Sesame Street' rejected as authoritarian The Times 7 Sept 1971
- Tizard B (1974) Preschool education in Great Britain Soc Science Rsh Council
- Tizard B (1976) Play: the child's way of learning? in Tizard B and Harvey B  
(Eds) Play in childhood Clinics in Developmental Medicine Spastics Int
- Tizard B, Burgess T et al (1980b) Fifteen thousand hours: a discussion  
Institute of Education U London

- Tizard B, Carmichael H et al (1980a) Four-year-olds talking to mothers and teachers in Hersov L and Berger M (Eds) Language and language disorders in childhood Pergamon
- Tizard B, Mortimore J and Burchell B (1981) Involving parents in nursery and infant schools Grant McIntyre
- Tizard B, Philips J and Plewis I (1976a) Play in pre-school centres - 1. Play measures and their relation to age, sex and IQ J C Psy Psychiatry 17 251-264
- Tizard B, Philips J and Plewis I (1976b) Play in pre-school centres - Effects on play of the child's social class and educational orientation of the centre J C Psy Psychiatry 17 265-274
- Tizard J (1974) Early malnutrition, growth and mental development in man Br Medical Bull 30 2 105-189
- Tizard J, Moss P and Perry J (1976) All our children Temple Smith/New Society
- Tizard J and Plewis I (1977) Critical notice: on Broman S H et al (1975) Pre-school IQ, prenatal and developmental correlates J C Psy Psychiatry 18 381-388
- Tizard J, Schofield W and Hewison J (1982) Collaboration between teachers and parents in assisting children's reading Br J Educ Psy 52 1-15
- Tough J (1977) How shall we educate the young child? in Davies A (Ed) Language and learning in early childhood Heinemann
- Toulmin S (1971) The concept of 'stages' in psychosocial development in Mischel T (Ed) Cognitive development and epistemology Academic
- Tukey J (1954) Causation, regression and path analysis in Kempthorne O et al (Eds) Statistics and mathematics in biology Iowa State College Press
- Tukey J (1962) The future of data analysis Annals of Mathl Statistics 33 1-67
- Tukey J (1970) Conclusions versus Decisions in Badia P. et al (Eds) Research problems in psychology Addison-Wesley
- Tukey J (1977) Exploratory data analysis Addison-Wesley
- Tversky A and Kahneman D (1971) Belief in the law of small numbers Psyl Bull 76 2 105-110
- Van de Geer J (1971) Introduction to multivariate analysis for the social sciences W H Freeman
- Van der Eyken W (1982) Home-Start: a four-year evaluation Home-Start Consultancy Leicester
- Vernon M (1957) Backwardness in reading: a study of its nature and origins Cambridge U P
- Vernon M (1969) Human motivation Cambridge U P
- Vernon P (1976) Environment and intelligence in Piaget, psychology and education Varma V and Williams P (Eds) Hodder and Stoughton
- Vernon P (1979) Intelligence: heredity and environment W H Freeman
- Versey J (1974) A longitudinal study of some aspects of cognitive development of primary school children Ph D thesis Institute of Education U London
- Vinod H D (1976a) Simulation and extension of a mean square error estimator in comparison with Stein's Technometrics 18 491-496
- Vinod H D (1976b) Application of new ridge regression methods to a study of Bell system scale economies J Am Stat Ass 71 836-841
- Wainer H (1976) Estimating coefficients in linear models Psyl Bull 83 213-217
- Wall W (1975) On education and equality Seminar Institute of Education U London (personal communication)
- Wall W and Williams H (1970) Longitudinal studies and the social sciences Soc Science Res Council / Heinemann
- Warlow A, Alexeev J et al (1977) Read! Some advice on how you can help your child Centre for Language in Primary Education, London
- Weber C, Foster P and Weikart D (1978) An economic analysis of the Ypsilanti Perry Preschool Project High/Scope Educ Res Foundation
- Wechsler D (1967) Wechsler Preschool and Primary Scale of Intelligence Psychological Corporation
- Weikart D et al (1970) Longitudinal results of the Ypsilanti Perry Preschool Project Office of Education Washington

- Weikart D, Bond J and McNeil J (1978) The Ypsilanti Perry Preschool Project: preschool years and longitudinal results through fourth grade High/Scope Educational Research Foundation
- Weikart D, Epstein A et al (1978) The Ypsilanti Preschool Curriculum Demonstration Project High/Scope Educational Research Foundation
- Wenar C (1963) The reliability of developmental histories Psychosomatic Medicine 25 6 505-509
- Werner E et al (1968) Prediction of intelligence and achievement at ten years from 20 month pediatric and psychological examinations C Dev 39 1063-1075
- Werts C, Jöreskog K and Linn R (1973) Identification and estimation in path analysis with unmeasured variables Am J Sociol 78 1469-1484
- Werts C and Linn R (1970) Path analysis: psychological examples Psyl Bull 74 3 193-212
- Werts C, Linn R and Jöreskog K (1974) Quantifying unmeasured variables in Blalock H (Ed) Measurement in the social sciences Macmillan
- White B (1977) Guidelines for parent education Planning Education Conference Flint Michigan
- White B, Kaban B and Attamucci J (1979) The origins of human competence: the final report of the Harvard Preschool Project Heath Lexington
- White B and Watts J (1973) Experience and environment: major influences on the development of the young child, Vol.1 Prentice-Hall
- Wichern D and Churchill G (1978) A comparison of ridge estimators Technometrics 20 301-311
- Widlake P (1975) The pre-school curriculum: a radical re-examination Education 3 to 13 April 1975 24-28
- Wilkinson E et al (1978, 1979) Strathclyde experiment in education: Govan Project: evaluation reports U Glasgow
- Willems E and Rausch H (1969) Naturalistic viewpoints in psychological research Holt, Rinehart and Winston
- Williams H L (1973) Compensatory education in the nursery school in Chazan M (Ed) Compensatory education Butterworth
- Winch R and Campbell D (1970) Proof? No. Evidence? Yes. The significance of tests of significance. in Morrison D and Henkel R (Eds) The significance controversy Butterworth
- Wiseman S (1956) The use of essays in selection at 11 plus: reliability and validity Br J Educ Psy 26 172-179
- Witelson S and Pallie W (1975) Left hemisphere specialisation for language in the newborn: neuratomic evidence for asymmetry Brain 96 641-646
- Witkin H et al (1962) Psychological differentiation Wiley
- Witty P (1950) Reading success and emotional adjustment Elementary English 27 5 281-296
- Wold H (1961) Unbiased predictors in Neyman J (Ed) Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability U California
- Wold H (1975) Path models with latent variables: the NIPALS approach in Blalock H et al (Eds) Quantitative sociology Academic
- Wolf R M (1966) Identification and measurement of environmental process variables related to intelligence Ph D thesis U Chicago
- Wolfson J (1976) Cultural influences on cognition and attainment in Wolfson J (Ed) Personality and learning Open University / Hodder and Stoughton
- Wood D et al (1980) Working with under fives Grant McIntyre
- Woodhead M (1976a) An experimental in nursery education National Foundation for Educ Resh
- Woodhead M (1976b) Intervening in disadvantage: a challenge for nursery education National Foundation for Educ Resh
- Woodhead M (1977) Pre-school education in Europe Committee for General and Technical Education Council of Europe
- Wragg E C et al (1980) The Butter Research Perspectives 1 U Exeter
- Wright H (1960a) Observational child study in Mussen P (Ed) Handbook of research methods in child development John Wiley
- Wright S (1921) Correlation and causation J Agricultural Resh 20 557-585

- Wright S (1960b) Path coefficients and path regressions: alternative or complementary concepts? Biometrics 16 189-202
- Wynn M (1972) Family policy Pelican
- Yarrow L, Klein R et al (1975) Cognitive and motivational development in early childhood in Friedlander B et al (Eds) Exceptional Infant, Vol.3: assessment and intervention Brunner/Mazel
- Yarrow M et al (1970) Reliability of maternal retrospection: a preliminary report Monog Socy for Rsh in C Dev Serial 138 35 5
- Yule W et al (1967) The reliability and validity of the Goodenough-Harris drawing test Br J Edul Psy 37 110-111
- Yule W, Rutter M et al (1974) Over- and under-achievement in reading: distribution in the general population Br J Edul Psy 44 1-12
- Zigler E (1972) Project Head Start: success or failure? Mimeograph Yale U
- Zimiles H (1970) Has evaluation failed compensatory education? in Hellmuth J (Ed) Disadvantaged child, Vol.3 Brunner/Mazel

Detailed Contents ...../

Detailed Contents

Index of Sections

Abstract	vi
Introduction	vii
1.00 Study Outline	1
2.00 Early Educational Attainment : a conceptual model	5
2.10 The development of early reading	7
2.11 Some reading theories	7
2.12 The debate on teaching methods	9
2.13 Environmental contributors to early reading	11
2.14 The prediction of early reading attainment	13
2.15 The case for early reading	14
2.20 The development of early mathematics	17
2.30 The influence of ability	22
2.31 Influences on ability	22
2.32 Damaged and delayed development	24
2.33 The link between ability and attainment	26
2.40 The influence of temperament	27
2.41 Early predictors of temperament	27
2.42 Characteristics of particular interest	28
2.43 Motivation	31
2.50 The influence of parents and home environments	34
2.51 The social milieu	34
2.52 Parent-child interaction	36
2.53 Child-rearing norms	39
2.54 Child-educating norms	40
2.55 The parent as teacher	43
2.60 The influence of pre-school environments	46
2.61 The history and provision of pre-schooling	46
2.62 The role and functions of pre-schooling	50
2.63 Research findings on nursery education	57
2.70 The influence of television	63
2.80 An integrated model	66
3.00 Intervening in Disadvantage : home and pre-school	72
3.10 Educational and social disadvantage	74
3.11 The home environment	74
3.12 The wider social environment	75
3.13 The school environment	79
3.20 The reality of intervention	82
3.21 Head Start: the American dream	83
3.22 Educational Priority Areas: the British compromise	86
3.23 Three landmark studies	88
3.24 Institutional intervention	92
3.25 Parent intervention	101

3.30	The methodology and research perspectives	108
3.31	Goal-setting and strategies	108
3.32	Design considerations	112
3.33	Sample considerations	114
3.34	Research perspectives	116
3.40	Evaluation, cost-benefit, dissemination and policy	120
3.41	Evaluation	120
3.42	Cost-benefit analysis	124
3.43	Dissemination and public policy	127
3.50	The present study: concepts and hypotheses	131
4.00	Intervention Project	134
4.10	Rationale	135
4.20	Sample	141
4.21	Sampling: concepts and limitations	141
4.22	Choice of sample	143
4.221	Sample population characteristics	143
4.222	Nursery facilities and the distribution of under-fives	144
4.223	Provision of books	146
4.224	Selection of sample schools and their children	146
4.225	School, class and staff characteristics	148
4.226	Selection of parents	150
4.30	Method	153
4.31	Child pre-tests and mid-tests	158
4.311	Pre-tests (nursery)	162
	Attainment tests	162
	Cognitive skills	167
	Needs assessments	176
4.312	Mid-tests (reception class)	177
4.32	Parent interviews	178
4.321	Areas of inquiry	180
4.322	Controls and scoring	182
4.33	Parent programmes	184
4.34	Child post-tests	196
5.00	Analysis	198
5.10	Statistical options	203
5.101	Descriptive analyses	207
5.102	Interpretive analyses	207
5.103	Particular techniques and their problems	209
5.11	Bivariate relationships	213
5.12	The multivariate approach	216
5.121	The component and factor hypotheses	217
5.122	The interpretation of variance	219
5.13	Causal models	222
5.14	Reliability and disattenuation	226
5.141	Reliability and validity of interviews	230
5.142	Disattenuation of the working matrix	232
5.15	Validity assessment	236
5.151	Rationale for a Redundancy Index	243
5.152	Examples of the Redundancy Index	246

5.16	The significance concept	251
5.17	Age and time concepts	254
5.171	Use of age variables	257
5.172	Use of time variables	258
5.173	Variables available for analysis	259
5.20	Multiple regression	260
5.201	Practical applications	261
5.202	Particular problems	262
5.203	Cross-validation	265
5.21	The limitations of Ordinary Least Squares	267
5.211	Statistical limitations of the regression method	268
5.212	Attempts to overcome the problem	269
5.22	V-ridge regression: theory and evidence	272
	Abstract	272
5.221	Ridge regression: its origins	273
5.222	Recent developments	275
5.223	Criticisms of ridge	277
5.224	The Vinod method	280
5.225	Derivation of IRSM and its corresponding k	282
5.23	Empirical assessment of the V-ridge model	285
5.231	Rationale underlying the development of the Vinod algorithm	287
5.232	Results	293
	Simulated intervention data	293
	Real intervention data	298
	Simulated mortality data	300
	Real mortality data	305
5.233	Conclusions	
5.234	Application of V-ridge to present study	314
5.30	Path analysis	318
5.301	Practical applications	319
5.302	Theoretical development	320
5.303	Specific approaches	322
5.304	Standardisation	327
5.305	Choice of algorithm	328
5.306	Disattenuation of all matrices	332
5.307	Data-handling programme	333
6.00	Results	335
6.10	Child tests	337
6.20	Parent interviews	344
6.21	Extraneous issues	344
6.22	Group allocation	345
6.23	Goals and ethical issues	346
6.24	Comments on nursery education	348
6.25	Reading activities	350
6.26	Mathematics activities	353
6.27	Attitude to school and school work	354
6.28	Television viewing	356
6.29	An ideal parent?	358

6.30...../

6.30	Parent programmes	360
6.31	'Selling' the programmes	360
6.32	Attendance pattern	362
6.33	The dropout phenomenon	365
6.34	Inherent problems	366
6.35	The programme meetings	369
6.36	Parents' views on the programmes	373
6.37	Some methodological considerations	378
6.38	Lessons for future programmes	381
6.40	Data, sub-sample and group characteristics	384
6.41	Broad characteristics of the data	385
6.411	Pre-test data (nursery)	388
6.412	Mid-test data (reception)	390
6.413	Post-test data	393
6.414	Parent-home interviews and programme scores	394
6.415	Other measures	395
6.416	Features of particular variables	395
6.42	Sub-sample characteristics	397
6.421	Morning and afternoon sub-samples	397
6.422	Male and female sub-samples	398
6.423	Sub-samples from advantaged and disadvantaged areas	398
6.424	Sub-samples from black and other groups	400
6.43	Comparison of parent group characteristics	405
6.431	Disadvantaged areas: working parents vs. non-working groups	409
6.432	Disadvantaged areas: working parents vs. programme attenders	409
6.433	Disadvantaged areas: programme attenders vs. non-attenders	410
6.434	Disadvantaged areas: reading attenders vs. mathematics attenders	411
6.435	All schools: 2-weekly reading attenders vs. 6-weekly reading attenders	412
6.44	The experimental design and analysis	413
6.50	Data: reliability and validity	419
6.51	Reliability determination	421
6.511	Pre-test reliabilities	421
6.512	Mid-test reliabilities	424
6.513	Post-test reliabilities	425
6.514	Parent interview reliabilities	426
6.515	Reliabilities of other data	430
6.516	Reliabilities of latent variables	432
6.52	Validity determination	433
6.521	Nomological validities of individual variables	437
	Nursery Ability variables	437
	Initial Attainment variables	440
	Parent Programme variables	441
	Parent Academic Environment variables	442
	Reception Ability variables	443
	Nursery Needs and Reception Needs	444
	Single variables	445
6.522	Nomological validity parameters for regression models	448
6.523	Specific validity assessments	452



6.60	Data: age and time variables	456
6.61	Age relationships	457
6.62	Time relationships	462
6.63	Derivation of new time variables	465
6.70	Correlations and group scores	470
6.71	Transformations of variables	473
6.72	Group scores	474
6.73	Predictive power of meetings attended	476
6.74	Correlational evidence on other variables	481
6.80	Path analyses	483
6.81	Application of regression techniques	484
6.811	V-ridge algorithm	484
6.812	Stepwise, hierarchical or simultaneous entry	486
6.813	Minor perturbation of non-invertable matrices	486
6.814	Covariance or correlation-based regression	487
6.815	Standardisation?	487
6.816	Squared multiple and single correlations	488
6.817	'Suppressor variables'	488
6.818	'Error' variance	490
6.82	Construction of path models	491
6.821	Basic design	492
6.822	Detailed construction	493
6.823	Limitations of the model	496
6.824	Procedures for inclusions and exclusion	496
6.825	Satellite models	501
6.826	Reliability of predictions	504
6.83	Specific criteria	505
6.831	Construction of the outcome variables	505
6.832	Limits set for probability and unique variance	507
6.833	Other criteria	510
6.84	Path Models 1 to 10	512
	Model 1: Total Disadvantaged Sample predicting Total Attainment	516
	Model 2: Advantaged Sample predicting Total Attainment	527
	Model 3: Disadvantaged Working Group predicting Total Attainment	539
	Model 4: Disadvantaged Reading Groups predicting Reading Attainment	549
	Models 5 and 6: Disadvantaged Maths Groups predicting Maths Numeracy Attainment and Maths Concepts Attainment	560
	Models 7 and 8: Disadvantaged Girls and Disadvantaged Boys predicting Total Attainment	576
	Models 9 and 10: Disadvantaged Black Children and Disadvantaged White Children predicting Total Attainment	589
6.90	Cost-benefit indicators	605
6.91	Comparison of costs	605
6.92	Comparison of benefits	609
6.93	Cost-effectiveness considerations	612

7.00	Synopsis	614
7.10	Summary of previous chapters	614
7.20	Comparison of variance predictions	659
7.30	Principal findings	665
7.40	The future?	672
References		674
Detailed Contents		
	Index of sections	695
	Index of tables	700
	Index of illustrations	701
	Index of appendices	702
Author Index		704
Subject Index		708
Appendices		R 1 to R 321

### Index of Tables

Table 1.	Summary of key points emphasised during programme meetings	188
Table 2.	Presentation of nomological validity within two typical models	241
Table 3.	Prediction of post-test reading attainment (disadvantaged reading groups)	246
Table 4.	Prediction of post-test total attainment (disadvantaged Black children)	247
Table 5.	Prediction of post-test total attainment (disadvantaged girls)	247
Table 6.	Prediction of post-test total attainment (disadvantaged boys)	248
Table 7.	Prediction of post-test total attainment (disadvantaged White children)	249
Table 8.	Correlation matrix and coefficients for simulation sample set A	295
Table 9.	Comparison of OLS and V-ridge regression on sample set A (7 regressor variables)	296
Table 10.	Comparison of coefficient stability across OLS and V-ridge solutions	297
Table 11.	Comparison of OLS and V-ridge regression on sample set B (6 regressors)	299
Table 12.	Correlation matrix and coefficients for simulation sample set C	301
Table 13.	Effect of error level: OLS vs. V-ridge in sample C (7 regressors)	303
Table 14.	Effect of variable sets: OLS vs. V-ridge variable sets C (N=100)	305
Table 15.	Basic parameters of sample set D (60 cases)	306
Table 16.	Comparison of OLS and V-ridge regression on sample set D (60 cases)	307- 308
Table 17.	Trial runs to test IVE improvement with different programme options	316
Table 18.	Mean data for attrition sample compared with a) main surviving sample; b) main surviving disadvantaged sample	339
Table 19.	Summary of main variables used in the analyses	386- 388
Table 20.	Parent groups: partitioning	406
Table 21.	Pre-test reliability data (test-retest correlations)	422

Table 22.	Mid-test reliability data (test-retest correlations)	424
Table 23.	Post-test reliability data (test-retest correlations)	425
Table 24.	Adjusted reliability values for parent interview variables	427
Table 25.	Validity of parent interview variables	428
Table 26.	Comparison of estimates of children's viewing time	429
Table 27.	Parent interview reliability estimates	430
Table 28.	Nomological validity comparisons for variables used in regression	435-436
Table 29.	Nomological validity parameters for regression models	449-451
Table 30.	Selected correlations between age, ability and attainment variables	459
Table 31.	Results of attainment regressions on cohort age and time in reception, using whole sample (N=159)	464
Table 32.	Selection of group mean scores and standard deviations for various samples	475
Table 33.	Correlations and regression parameters for Programme meetings	477
Table 34.	Squared correlations between parent attendance and child performance	479
Table 35.	Comparison of regression solutions for a small sample (N=28)	485
Table 36.	Comparison of probability and unique variances in a path regression	497
Table 37.	Probability and unique variances in reduced path regression	498
Table 38.	Probability and unique variance limits for path models	508
Table 39.	Squared correlations between E.P.V.T. and Ability and Attainment	534
Table 40.	Cost in equivalent teacher hours of three intervention programmes	609
Table 41.	Hypothesised effectiveness of three intervention programmes	611
Table 42.	Comparisons of the cost-effectiveness of three intervention programmes	658

### Index of Illustrations

Figure 1.	Conceptual models of early mathematical development	21
Figure 2.	Theoretical model of contributors to early educational attainment	70
Figure 3.	Research schedule	157
Figure 4.	Alphabet word card - reading programme	189
Figure 5.	Pages from second reader - reading programme	190
Figure 6.	Typical number/symbol cards - mathematics programme	191
Figure 7.	Set identification and matching activities (parent to read and explain instructions) - mathematics programme	192
Figure 8.	The ridge trace: examples of typical alterations in coefficient values	276
Figure 9.	Plot of ridge coefficients against Vinod's m	281
Figure 10.	Simple recursive path analysis model	318
Figure 11.	A non-recursive path analysis model	329
Figure 12.	Application of the NIPALS method to a path model	330
Figure 13.	Frequency distribution of reading awareness (Nursery)	389
Figure 14.	Frequency distribution of WPPSI Information (Reception)	391
Figure 15.	Block design histograms and the nursery and reception levels	392
Figure 16.	Item from the WPPSI Picture Completion Sub-test	453
Figure 17.	Histograms of Nursery age, Reception (mid-test) and Post-Test age	458
Figure 18.	Histograms of raw cohort age, time in nursery and time in reception	463

Figure 19. Diagrammatic representation of cohort age and time concepts	466,
	646
Figure 20. Histograms of composite Time in Nursery and Time in Reception	468
Figure 21. Plot of relationships attendance at Programme meetings	471
Creation of latent variables for model 1	517-518
Path model 1	521
Creation of latent variables for model 2	528-529
Path model 2	531
Creation of latent variables for model 3	540-541
Path model 3	543
Creation of latent variables for model 4	550-551
Path model 4	553
Creation of latent variables for models 5 and 6	561-564
Path models 5 and 6	567-568
Creation of latent variables for models 7 and 8	577-580
Path models 7 and 8	583-584
Creation of latent variables for models 9 and 10	590-593
Path models 9 and 10	596-597
Figure 22. Comparison of variance predictions across path models: Disadvantaged Sample vs. Advantaged Sample, predicting Total Attainment	661
Figure 23. Comparison of variance predictions across path models: Disadvantaged Reading Groups predicting Reading vs. Disadvantaged Maths Groups predicting Maths Concepts	662
Figure 24. Comparison of variance predictions across path models: Disadvantaged Girls vs. Disadvantaged Boys, predicting Total Attainment	663
Figure 25. Comparison of variance predictions across path models: Disadvantaged Black Children vs. Disadvantaged White Children, predicting Total Attainment	664

### Index of Appendices

A. Test descriptions and scoring	
A1. Summary of test appendices	R 3
A2. Pre-test attainments	R 4
A3. Pre-test cognitive performance	R 8
A4. Mid-test cognitive performance	R 15
A5. Post-test attainments	R 16
A6. Needs assessments	R 17
A7. Additional mathematical concept pictures	R 19
A8. Matching Familiar Figures pictures	R 20
A9. Patterns for visuo-motor integration test	R 35
A10. Modified Goodenough-Harris scoring protocol	R 36
A11. Sample self-pictures	R 37
A12. Multiplicative classification pictures	R 41
B. Interviewing	
B1. Interview wording	R 42
B2. Interview schedule	R 48
B3. Interview coding	R 51
B4. Preliminary letter to parents	R 55
C. Parent programmes	
C1. Summary of parent programme documents	R 56
C2. Basic documents relating to programmes	R 57
C3. Reading meetings - agendas	R 64
C4. Reading programme - ancillary materials	R 71

C5.	Mathematics meetings - agendas	R 106
C6.	Mathematics programme - ancillary materials	R 113
D.	Data and main parameters	
D1.	Summary of statistical appendices	R 138
D2.	Raw data	R 139
D3.	Basic distributional characteristics	R 156
D4.	Selected histograms	R 158
D5.	Problem features of particular variables	R 168
D6.	Reliabilities of variables	R 174
D7.	Nomological validity parameters	R 175
D8.	Sub-sample characteristics	R 196
D9.	Parent group characteristics	R 199
D10.	Validity of rhythmic ability and distractibility assessments	R 202
E.	Computer programmes	
E1.	V-ridge regression programme	R 203
E2.	V-ridge programme: user's manual	R 278
E3.	Typical plots of log ISRM vs. log k (V-ridge)	R 304
E4.	Remul file-handling programme	R 308
E5.	Remul programme: user's manual	R 314

Author index ...../

Author Index

Only the names of first authors are cited here

Abt Associates	85 127 128	Blackstone T	102 624	Cane B	10
	204	Blackstone W	20	Cannell C	178 231 416
Achenbach T	168	Blalock H	214 223 320		419 428
Acland H	87 88 622		322 327	Central Policy	
Ainsworth M	37	Blank M	98 138 623	Review Staff	35 60
Ambron S	116 117 626	Block J	168	Chall J	9
Annon P	171	Bloom B	232	Chalmers C	293
Anastasi A	171	Blum M	178 231 428	Chamberlain R	67 76
Anderson L	126	Bock R	207 234 235	Chazan M	57 58 61
Anderson R	85 127 199	Boehm A	166		79 80 93
	204	Bohrnstedt G	227 228		96 98 123
Andersen S	340	Bond G	10		136 139 159
Andison F	63	Bondon R	322		619 620 621
Anscombe F	210	Bougere M	162	Cicirelle V	84 212
Applebee A	198	Bowlby J	36	Clark M	44
Armstrong G	87	Boyle R	216	Clarke A	24 84 110
Athey I	29	Brainerd C	506	Clarke W	18 165
Atkinson D	52	Brandt R	158	Clay M	165
Ault E	168	Brimer M	162 239 422	Clements M	20
Austin G	46 48 83		452	Clifford C	30
	84 97 618	Broman S	35 67 270	Cochrane K	164
		Bronfenbrenner U	39 54	Cohen J	209 212 232
			75 82 110		233 260 264
Badia P	158		111 122 128		269 302
Bakan D	251		129 136 403	Coleman J	41 68 75
Bakker D	167		617 619 621		81 261 262
Baltes P	256		625 628		319 320 324
Banks O	74	Brown F	227 229 238		472
Banta T	159	Brown P	277	Comber L	75
Barker R	158	Bruner J	23 31 37	Coniffe D	277
Barker W	33 130 135		60 96 616	Cookey W	217 268
	154 159 163	Bryant B	50 60	Costner H	322
	164 196 199	Bryant P	171	Council of Europe	61
	262 272 393	Buffery A	20	Cowles M	201
	401 423 454	Bthler C	32 346	Cox A	178 179 427
	455	Bullock Report	15 42 58	Crane W	154
Bartke J	229		98	Craviato J	23
Baumrind D	32	Burroughs G	141 142 211	Cronbach L	113 153 158
Beaton A	268	Busch J	64		175 209 213
Becker G	126 127	Butcher H	22		216 226 228
Becker W	31 99 623	Butts R	83		232 237 238
Bell L	179				431 634
Bennett N	128 627			Crowe B	51
Bercovici A	39	Cain G	261	Cunningham C	137
Bereiter C	32 94	Cairns E	423	Cureton E	218
Berg I	83	Caldwell B	23 75 103		
Bernstein B	12 138		624		
Bibby J	261 269	Campbell D	85 204 205	Daniels J	454
Billewicz W	142		209 211 212	Darlington R	264 489
Birch H	167		232 236 237	Davidson R	160
Bissell J	83 94 95		262 265	Davie R	67 75 76
					176 179

Day M	116			Fowler W	203			Havlieck L	213		
Deasey D	47	83		Francis H	208	209		Hays W	207		
Deegan J	268			Fuller W	234			Heath A	80		
De Finetti B	210							Heber R	24	88	116
De Hirsch K	13	167	170						122	124	622
	552			Gallagher J	321			Hewison J	43	91	178
Dempster A	277			Galtung J	182	183			204	355	613
Denney D	29	168		Gammage P	30				618	622	
Dept. Edu Science	93	96		Garber H	88	203	205	Hindley C	209	224	256
Detheux M	100			Gardner P	183				321		
Deutsch M	25	182		Gaudy E	31			Hochberg J	9		
De Vries A	93	203		Gehlbach R	58			Hocking R	277		
Dimitrovsky L	172			Gewirtz J	255			Hoerl A	269	273	274
Dobbing J	22			Ghikas C	65	357		Holland S	93		
Doman G	44			Gibbons D	278			Holmes D	9		
Donachy W	105			Gibson E	9			Holt J	135		
Douglas J	41	58	67	Gocka E	201			Horowitz F	122		
	75	76	179	Goldberg A	223	319		House E	85	127	128
	181	182	346	Goldberg M	123				199	204	419
	617			Goldberger A	320	322	323	Hughes M	35		
Downing J	10	12	15	Goldstein H	200	234	242	Hull C	31		
	163	164			254	255	260	Hunt J McV	25	153	615
					261	319			625		
Drabman R	161			Goldstein M	275			Hutchison D	13		
Draper N	260	278	279	Golub G	265			Hutt C	20	23	158
Duncan O	233	319	320	Goodacre E	10	15			587		
	322	323	324	Goodenough F	170	423		Hyman H	427		
Dukes W	201			Goodman L	322						
Durdin-Smith Comm.	63	357		Gordon I	104	117	624				
Durkin D	14	15	29	Gordon R	261	264	268	Igaga J	642		
	164	615			331			Illich I	135		
Dwyer C	12			Gorsuch R	271						
Dweck C	30			Gottman J	201						
				Gourlay N	212			Jencks C	77	126	262
Eaves L	153	470		Gray S	113	114	203		324	325	472
Edmonds R	325			Green B	216			Jensen A	122	400	472
Edwards A	208	245		Green R	17				599	602	
Ehrenberg A	141			Grieger P	80			Johnson D	104		
Elardo R	178			Groff P	167			Johnson S	158		
Elkind D	31			Guilkey D	275			Johnston J	200	323	
Elkonin D	9	11	615	Gulliford R	162			Jöreskog K	324		
Ellenberg J	210			Gumbert E	83						
Evans R	199			Gunst R	275						
Eysenck H	27	95	141	Gutelius M	203			Kagan J	29	159	168
									400		
Farebrother R	277			Haavelmo T	323			Kamii C	154	155	
Farnham-Diggory S	8			Halliday M	37			Karnes M	94	104	203
Farr R	164			Halsey A	77	81	87		403		
Farran D	38	617			108	109	319	Keddie N	78		
Feiler A	8				320	325	621	Kellaghan T	93	179	204
Ferguson G	226	229			625			Kelsall R	76	83	
Ferguson N	88			Hamill D	8			Kendall M	214	216	261
Ferri E	51	106							264	278	
Finn J	217			Hardy J	173			Kerlinger F	203	213	237
Flavell J	171	172	506	Harris C	209				260	262	264
Floud J	80			Harvard Bdul Rev	77				269	311	
Fogelman K	209			Hauser R	319	328	330	Kessen W	140		
					488	491		Kim J	216		

King W	237	Marwell A	173 219	Poulton G	111
Kish L	211	Mayall B	50	Price B	277
Kitay P	169	Meehl P	251	Pring R	633
Klein L	323 330	Miller L	116 626	Pumfrey	162 163 196
Koppitz E	169 208 239	Mischel W	27		
	423	Modgil S	171		
Kosc L	20	Moore T	256	Quigley H	100
Koutsoyiannis A	323 324	Morris A	106		
Krutelskii V	165	Morrison D	252 253		
		Mortimore J	88	Ramey C	69 95
Labov W	138	Mosbaeck E	328	Raven J	105
Laing A	53	Moseley C	162	Reid J	163 187
Lambie D	89	Mosteller F	81 267	Ribich T	126
Land K	321	Mpiangu B	19 172	Richards M	37
Landsdown R	20			Richardson K	179
Lashley K	167	Namboordiri N	205 206 216	Richardson S	28
Lawley D	218 260 270		217 260 320	Rose N	51 106
Lawrence D	29		326 327 329	Rosen B	32
Lazar I	85 109 115		484 500 507	Rosen H	138
	118 204 472		631	Rosenfield A	162
	622	Narrol H	153	Rosenthal M	38
Leach P	35	Neligan G	199	Rosenthal R	80 201 203
Leitenberg H	201	New Society	34	Rosetti Ferreira M	38
Lesser H	64 107 357	Newsom Report	75 76	Rothenberg J	126
Levenstein P	104	Newsom J	37 40 178	Rozeboom W	278
Levin H	126		179 181 346	Runkel P	142
Lindeman R	237	Noonan R	68 328 330	Rutter M	30 31 37
Lindley D	208		331 494 505		68 76 80
Lines R	18 87 97	Novick M	265		128 621 627
	165	Nunnally J	251		
Lister I	12 135			Sandersen A	14
Little A	403	Obenchain R	275	Santa J	171
Lohnes P	166	Olson D	175	Sattler J	173 391 452
Lord F	209	Osterreith P	93	Schaefer E	40 136
Lovell K	18 19 163	Overton E	144	Schaffer H	37
	165 172 193	Owen G	101	Schaie K	256
Lunzer E	271			Schmidt F	271
				Schönemann P	218
McCall R	216	Pantin G	107	Schraum W	63
Maccoby E	231	Parry H	428	Schultz C	32
McDonald G	277 278 293	Pasmanick B	28	Seaver W	80
	300 302 305	Peaker G	68 75 109	Seligman M	30
	311		111 141 261	Sharot T	210
Macdonald K	321 326 327		319 324	Shearer D	107
Mackay D	11 44	Perney J	64 357	Shipman V	83
McKay H	95 96	Phares M	30	Sigel I	159 171 342
McKennell A	233	Piaget J	23	Silverstein A	174
McMillan J	44 186	Pidgeon D	237	Simon H	213 256
McNemar Q	201 205 262	Plowden B	51 619 620	Sinclair H	165
	482 488	Plowden Report	48 51 68	Skatkin L	19
Madrid-Jimenez R	178		75 86 87	Skeels H	103 624
Malipant R	9		319 618 619	Slater M	252
Marquardt D	277	Plutchik R	141	Snilansky M	117 118 121
Maslow A	32 176	Pond C	34		270 626
Matthews G	165	Pepper K	vii 222	Smith F	7 8 9
Mattingly I	8 10	Posner M	8		11 187 393
					614



Smith G	87 128	Vernon P	22 153
Smith J	179	Versey J	172
Smith K	164	Vinod H	vi 279 280
Smith M	104		281 to 288
Smith T	60		312 483 636
Southgate V	15 425 454 615		
Sovell T	599		
Spearman C	234	Wainer H	271
Spencer M	11	Wall W	74 321
Stahl D	165	Warlow A	359
Steffe L	18	Weber C	125 126 627
Stevens G	16	Wechsler D	166 239 424
Stevenson H	13 271	Weikart D	89 90 91
Stone M	265 290		111 622
Stott D	32	Wenar C	231
Strauss S	171	Werts C	216 324 327
Streissguth A	39	White B	22 111 122
Stukát K	60		182 346 625
Sullivan J	322	Wichern D	275
Swan D	26	Widlake P	60
Swenson E	165	Wilkinson E	203
Sylva K	60	Williams H	100 623
		Winch R	252
		Wiseman S	229 230
Tanner J	171 255	Witelson S	17
Taruntaeva T	193	Witkin H	29
Taylor M	600 601 602	Witty P	30 31
Taylor P	52	Weld H	261 328 329
Tew B	8		330 331 483
Thackray D	14 164		505
Theil H	207 208 260	Wolf R	32 182
	323 324	Wolfson J	76
Thisted R	288	Wood D	60
Thomas A	27 178 179	Woodhead M	42 52 100
Thorndike R	75		211 619 623
Times, The	64	Wragg E	80 633
Tizard B	42 53 54	Wright H	158
	57 58 59	Wright S	320
	60 80 91	Wynn M	36 616
	93 102 105		
	111 138 348		
	447 617 618	Yarrow L	24
	619 624	Yarrow M	231 427
Tizard J	22 43	Yule W	26 167 170
	150 270 601		214
	618		
Tough J	53 98		
Toulmin S	171	Zigler E	83
Tukey J	198 210 213	Zimiles H	123
	251 318 320		
	470 632		
Tversky A	142 201		
Van de Geer J	219 264 271		
Van der Byken W	106		
Vernon M	31 32 167		
	182 346 403		

Subject Index...../

Subject Index

Usually only the first page number in any brief sequence is noted

- |                         |                                    |  |
|-------------------------|------------------------------------|--|
| ability                 | assessment                         | 159 173 399 401 437 Appendix A3 A4     |
|                         | general                            | 22 26                                  |
|                         | meta-cognitive                     |  |
|                         | skill                              | 167 Appendix A8 A9 A10 A11             |
|                         | reading link                       | 26                                     |
| academic attainment     | <u>see</u> attainment              |  |
| action research         |                                    | 108 121 625                            |
| additive validity       | <u>see</u> validity, nomological   |  |
| advantaged sample       | path model                         | 530 537 651                            |
| age factors             |                                    | 159 254 257 259 456 457 536 635 645    |
| analysis                | cross-validation                   | <u>see</u> cross-validation            |
|                         | intervention studies               | 203                                    |
|                         | longitudinal                       | 201 321                                |
|                         | principles                         | 198 205 413 472 631                    |
|                         | problems                           | 89 123 129 199 224 232 413 417 459 501 |
|                         | sensitivity                        | 205 216 419 485                        |
| analysis of variance    |                                    | 220 643                                |
| Asian                   | <u>see</u> ethnic issues           |  |
| assessment              | <u>see</u> ..... assessment        |  |
|                         | <u>see</u> evaluation              |  |
| attainment              | early                              | 5                                      |
|                         | failure                            | 25 30                                  |
|                         | general                            | 26                                     |
|                         | model                              | 66                                     |
|                         | prediction                         | 13                                     |
|                         | tests                              | 159 162 Appendix A2 A5                 |
| attention               | <u>see</u> distractibility         |  |
| attrition               | <u>see</u> sampling, attrition     |  |
| author index            |                                    | 704                                    |
| Bender Gestalt          | <u>see</u> ability, meta-cognitive |  |
| Block Design            | <u>see</u> Wechsler                |  |
| bivariate relationships |                                    | 213 470 481 632                        |
| Black (ethnic)          | <u>see</u> ethnic issues           |  |
| books                   |                                    | 146 350 628                            |
|                         | <u>see also</u> libraries          |  |
| Bronfenbrenner          |                                    | 39 54 82 111 128                       |
| causation               |                                    | 213 222 239 267 633                    |
| change scores           |                                    | 209                                    |

child minding		35 49
child rearing	<u>see</u> parent.....	
Coleman Report		81
cognitive development	<u>see</u> ability	
cognitive skills	<u>see</u> ability	
cognitive style		29
cohort concepts		256
communication	<u>see</u> language	
community environment		12 25 34 75
compensatory education		74 117
control groups	principle	211
	sample	<u>see</u> sampling, control groups
correlations	<u>see</u> bivariate relationships	
cost-benefit analysis		124 605 612 627 657
counselling of mothers		150
cross-validation		256 265 285 290
data	distributional	
	characteristics	388 457 463 468 473 641 Appendix D3 D4
	handling	333 Appendix E4 E5
	mid-test	390 424 443
	parent interview	394 426 442
	parent programmes	395 441
	pre-test	388 421 437 440
	post-test	393 425
	problems	395 459 Appendix D5
	raw	Appendix D2
	sample	<u>see</u> sampling, parent programme groups
		<u>see</u> sampling, sub-samples
day nursery care		35 50
design	research	108 112 153 199 413 625 629 630 643
development concepts		22 255 615
disattenuation		226 231 232 288 332 633
disadvantage	general	25 74 78
	<u>see also</u> social class.....	
disadvantaged sample	path model	516 525 649
dissemination		127 627
distractibility		29 176 452
draw-a-person	<u>see</u> self-picture test	
early attainment	<u>see</u> attainment	
early experience		24
Educational Priority Areas		18 86 109 115 622
English Picture Vocabulary Test		162 446 473 489 523 533 555

environmental enrichment		47 100
equality	<u>see</u> compensatory education <u>see</u> disadvantage <u>see</u> intervention <u>see</u> social class.....	
ethical issues		131 637
ethnic issues		77 347 400 474 599 600 602 603 642 656 660
ethnic differences	path model	589 603
European	<u>see</u> ethnic issues	
evaluation	criteria general test usage <u>see also</u> analysis..... <u>see also</u> parent programmes..... <u>see also</u> .....assessment	120 153 120 123 128 626 155 159
factor analysis		217 270
family	finances policy	36 36
Follow Through programmes		85 127 199
future possibilities		672
generalisability	<u>see</u> validity	
Head Start		83 115 621
health visitors		101
home environment		74 399 400 442 519 522 530 532 544 554 566 570 585 594 604 <u>see also</u> parent.....
home visiting		39 102 617
hypotheses	<u>see</u> research, hypotheses	
illustrations index		701
inequality	<u>see</u> social class.....	
Infant Reading Test		154 163 388 393 455 Appendix A2
intelligence	general <u>see also</u> ability	22 121
integration of skills		169 538
intervention	effectiveness evaluation general process <u>see also</u> parent programmes.....	117 601 <u>see</u> evaluation..... 72 82 92 621 622 623 624 116 123 626
interviews	ethical issues parent protocol	346 178 182 344 394 637 Appendix B1 B2 B3 B4

- jack-knife            see outliers
- language            assessment            162  
                     early development    37 138 534 555  
                     programmes            97 138 623  
                     see also English Picture Vocabulary Test
- libraries            146 352
- matching groups            210
- Matching Familiar Figures            168
- mathematics            assessment            165 506 Appendix A7  
                     early            17 21 353 615  
                     methods            17  
                     programmes general    97  
                     programmes, path model    560 574 654  
                     theories            17  
                     thinking            19  
                     see also parent programmes, mathematics
- mid-test data            see data, mid-test
- Milwaukee Project            88 110 116 122 124
- models            analysis            216 222 620  
                     validity            see validity, redundancy index
- mothers            employment            34 185 616  
                     role            35  
                     see also working group parents
- motivation            31 616
- multi-causal            see causation  
                     see multivariate
- multiple regression    alternative estimators    271 275  
                     capitalisation on error  
                     (variance estimate)    262  
                     general            200 220 260 470 635  
                     hierarchical            270 486  
                     multicollinearity    see problems  
                     Ordinary Least Squares    267 277 497 635  
                     perturbation            486  
                     practical applications    261 270 277  
                     problems            261 262 264 267 277 280  
                     reliability of  
                     coefficients            268 272 287 290  
                     ridge            see V-ridge  
                     stepwise            263 269 486  
                     suppressor variables    264  
                     V-ridge application    314 484 487  
                     V-ridge development    272 287 635 Appendix E1 E2 E3  
                     V-ridge evaluation    285 293 311 484 497  
                     V-ridge theory            272 277 280 282
- multivariate approach            113 153 198 205 216 255 472 632
- needs assessment            176 431 444 481 594 602 603 Appendix A6
- nomological validity    see validity, nomological

nursery assistants	100
nursery care	<u>see</u> day care
nursery data	<u>see</u> data, pre-test
nursery schooling	<u>see</u> pre-schooling
nutrition	23 28 38 49 96
outliers	210 264 271
parent	attitude towards school 354 638 behaviours 178 399 401 child interaction 36 366 617 child discipline 355 637 child-rearing 39 103 617 639 educational role 11 32 39 41 43 187 358 522 598 617 618 intervention 89 101 114 118 131 interviews <u>see</u> interviews, parent mathematical development of children 353 638 pressures on children 367 reading to children 350 637 638 reading, teaching of 11 351 socio-educational differences <u>see</u> social class, differences <u>see also</u> home environment
parent programmes	attendance and dropout 362 365 376 400 402 406 410 431 639 design 149 184 378 effectiveness 476 477 479 535 536 556 557 571 573 660 evaluation 380 476 frequency 375 goals 138 186 mathematics 193 373 411 476 Appendix C5 C6 parent groups <u>see</u> sampling, parent groups parents views on 373 640 predictive power 476 535 556 571 581 647 principles 122 188 381 581 641 problems 365 366 367 368 377 379 639 640 procedures and schedule 157 184 360 381 630 Appendix C1 C2 randomisation 147 152 rationale 136 reading 187 369 411 476 Appendix C3 C4 resources 186
path analysis	application 328 483 491 492 493 conceptual groups 493 criteria 496 507 510 649 general 11 257 318 636 648 latent variables 494 legend 513 outcome variables 505 procedures 496 505 648 satellite models 491 501 539 576 649 status variables 494 study models 483 491 <u>512</u> theory 320 322 326 328

reliability	data	385 419 421 430 432	Appendix D6
	general	226 633 643	
	interviews	230	
	test	229 421 425	
research design	<u>see</u> design, research		
research	findings	665	
	hypotheses	133 628 665	
rhythmic ability		167 401 437 443	
sampling	attrition	338	
	control groups	114 151 406 407 408 413 416	
	general	114 141 626	
	parent groups	405 411 417 643	Appendix D9
	procedures	146 150 341 345 379 413 629	
	size	96 142 201 267 406 414	
	sub-samples	397 400 474 641	Appendix D8
	<u>see also</u> population.....		
satellite models	<u>see</u> path analysis, satellite		
school	attainment influence	41 79 621	
	characteristics	148	
	intervention with		
	parents	43 91	
	parent involvement	11 42 60 74 149 344	
self-concept		30 176	
self-picture test		170 519	
sex differences	general	398 660	
	path model	576 587 655	
siblings		351	
significance concept		251 500 634	
	<u>see also</u> probability		
social class	differences	35 41 77 181 390 397 398 530 533 534	
		544 660	
	influence	99 130 616 620 623	
social policy		128	
social problems		150	
standardisation of data		327 487	
statistical options		203 207	
study programme		131	
	<u>see also</u> parent programmes.....		
structure in learning		23 53 94 98 378	
suppressor variables		264 488 648	
tables, index		700	
teachers	expectations	80	
	views	100	
television	viewing and effects	63 356 620 638	
temperament and attainment		27 616	

tests	choice	158	Appendices A
	interaction with		
	children	160 342 421	
	procedure and schedule	337 341 636	
	usage	155 159 254	
time	concepts	254 462 635	
	factors	258 456 465 645 646	
	in nursery and		
	reception	258 430 445 462 465 524 535 546 556	
		572 586	
Tizard, Barbara		53 102	
validity	general	236 634 644	
	interview	231 428	
	nomological	238 433 437 448 634 644	Appendix D7
	redundancy index	243 448 645	
	test	229 237 452	Appendix D10
variables		259 386	
variance	interpretation	219 244 488 490 504 632	
	prediction comparison	659	
V-ridge	<u>see</u> multiple regression, V-ridge		
Wechsler Pre-school and Primary Scale of Intelligence		173 390 452	
West Indian	<u>see</u> ethnic issues		
White (ethnic)	<u>see</u> ethnic issues		
working group parents	general	406 409 471	
	path model	539 547 652	
Ypsilanti Perry Pre-School Project		89 125	